

Review article

A review of pollution-based real-time modelling and control for sewage systems

Rodrigo da Silva Gesser^{a,b,*}, Holger Voos^a, Alex Cornelissen^b, Georges Schutz^b

^a University of Luxembourg, 29 Av. John F. Kennedy, Luxembourg City, 1855, Luxembourg

^b RTC4Water, 62a Grand-Rue, Roeser, 3394, Luxembourg

ARTICLE INFO

Keywords:

Real-time control
Water quality
Sewage systems

ABSTRACT

Conventional solutions for wastewater collection focus on reducing overflow events in the sewage network, which can be achieved by adapting sewer infrastructure or, a more cost-effective alternative, by implementing a non-engineering management solution. The state-of-the-art solution is centered on Real-Time Control (RTC), which is already resulting in a positive impact on the environment by decreasing the volume of wastewater being discharged into receiving waters. Researchers have been continuing efforts towards upgrading RTC solutions for sewage systems and a new approach, although rudimentary, was introduced in 1997, known as Pollution-based RTC (P-RTC), which added water quality (concentration or load) information explicitly within the RTC algorithm. Formally, P-RTC is encompassed of several control methodologies using a measurement or estimation of the concentration (i.e. COD or ammonia) of the sewage throughout the network. The use of P-RTC can result in a better control performance with a reduction in concentration of overflowing wastewater observed associated with an increase of concentration of sewage arriving at the Wastewater Treatment Plant (WWTP). The literature revealed that P-RTC can be differentiated by: (1) implementation method; (2) how water quality is incorporated, and (3) overall control objectives. Additionally, this paper evaluates the hydrological models used for P-RTC. The objective of this paper is to compile relevant research in pollution-based modelling and real-time control of sewage systems, explaining the general concepts within each P-RTC category and their differences.

1. Introduction

Modernisation increased the use of many fundamental resources and thus required an evolution of the techniques responsible for handling, distributing, collecting and disposing of these resources. The prime example is water, which is arguable the most important available resource, and wastewater, both having critical environmental, social and economical impacts [7]. In most countries, water is mainly collected, treated and distributed for personal use in dwellings, while wastewater is the result of its consumption: wastewater, is transported by the sewage system, treated in Waste Water Treatment Plants (WWTPs), and discharged into receiving water bodies such as river or seas [71]. Wastewater, the focus of this paper, needs to be treated to avoid negative impacts on the environment and, therefore, requires advanced techniques to guarantee that these resources are being efficiently and reliably collected [48].

* Corresponding author.

E-mail address: rodrigo.dasilvagesser@uni.lu (R. da Silva Gesser).

<https://doi.org/10.1016/j.heliyon.2024.e31831>

Received 7 November 2023; Received in revised form 2 May 2024; Accepted 22 May 2024

Available online 31 May 2024

2405-8440/© 2024 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

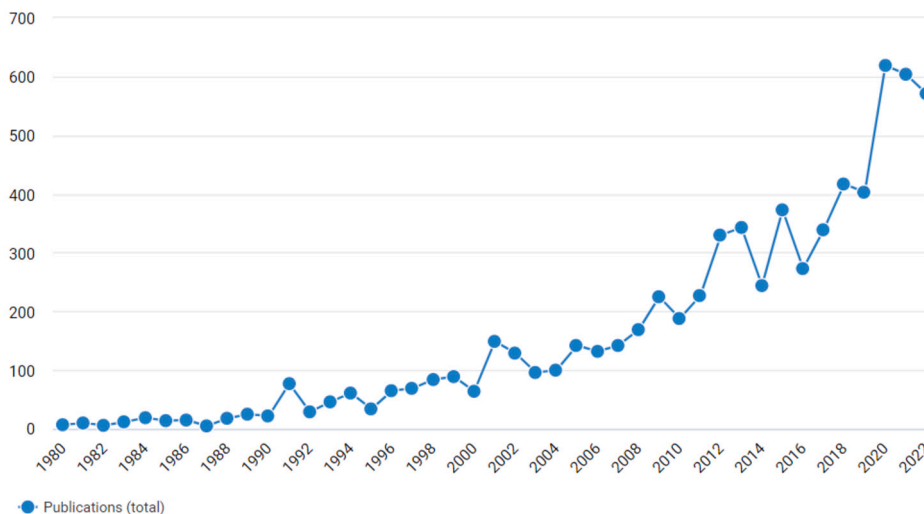


Fig. 1. Total publications per year from 1980 to 2022, estimated by searching the key words “Real time control” and “Wastewater” in the data-set provided by [34].

The main issue in sewage systems is the occurrence of overflows in a combined network, characterised by the combination of rainwater and sewage that are collected and transported through the wastewater system. Here, Combined Sewer Overflow (CSO) are caused by rainwater that flows into the sewage system during rain events and overloads the sewers due to insufficient storage [62]. Increases in CSO events may be caused by climate change, increases in urban population and reduction in soil rainwater retention [32]. For completeness, the papers [60] and [46] details the state of the art for overflow prediction in wastewater systems.

Therefore, minimising CSO events is of utmost importance for every urban area [62]. This can be achieved by increasing storage volume, resizing pipes and pumps or removing rainwater from the sewage system by adding Sustainable Urban Drainage Systems (SUDS). However, a more cost-effective solution is to implement software-based approaches. A few of such solutions have been implemented in actual applications, such as the study from [13] implementing a Model Predictive Control (MPC) strategy in Barcelona. [61] developed a global optimal control solution in Quebec. [25] and [26] implemented a multi-objective MPC approach in Luxembourg and [51] designed a rule-based algorithm to control a sewer system in Copenhagen.

Real-Time Control (RTC) is a common approach for global control of wastewater networks. The underlying reason for applying a RTC is that wastewater networks are systems with a highly dynamic behaviour, with constantly changing operational variables, necessitating a controller capable of reacting in real time [20]. The overall literature of RTC applied to drainage systems is growing constantly, as can be seen in Fig. 1 and according to the online data-set available at [34], the estimation of published studies on the topic is approximately 7000 papers. There are several reviews, detailing many different algorithms and implementations, for example: [32], [73] and [95] reviewed modelling and real-time control of sewage systems, in the research of [16] the performance of different RTC methods is evaluated and [96] illustrated the impact of RTC algorithms on storm-water control measures.

However, the usual application of RTC does not implicitly consider the pollution load [80], which is only feasible by either using more complex models or by adding water quality information to the algorithm. Measuring water quality in the sewage system is difficult because of the harsh environment and associated maintenance costs [19]. Incorporating quality data results in the possibility to explicitly influence the sewage network and manipulate the system to ensure better performance, which is possible, but not guaranteed using a purely volume-based approach. Improvements derived from using water quality data in RTC can include reducing overflow with highly concentrated wastewater and increasing load being directed to the WWTP.

Therefore, efforts have been made to improve RTC using pollution-based approaches as seen in the works of [48], [78] and [85]. Also, studies in water quality prediction and assessment in sewer systems are available [39], [57], [99].

Although in the literature there are reviews on water quality modelling [100], water quality dynamics and in-sewer processes [31], a search through SCOPUS and Google Scholar revealed a lack of studies reviewing specifically Pollution-based RTC (P-RTC).

Therefore, the aim of this paper is to provide a review of P-RTC, focusing on the models used to describe pollution in SUDS for control purposes and related control techniques that have been developed so far. It is imperative to provide a comprehensive guide for researchers and practitioners about the substantial impact of applying RTC with explicit quality objectives. Finally, the paper discusses what new research could result from the identified literature gaps and how filling these gaps might influence the future of sewage systems control and modelling.

2. Modelling of sewage systems for control

The sewage system infrastructure is designed to cope with the inflow of wastewater originating from the daily use of water discarded from households and industry, which is defined as the Dry Weather (DW) inflow. However, during rain events, a secondary inflow enters the sewer system, also known as Wet Weather (WW) inflow, and the capacity to store the volume of water might exceed the maximum storage volume of a retention structure (pipe or tank). As can be seen in Fig. 2, this then results in a CSO event.

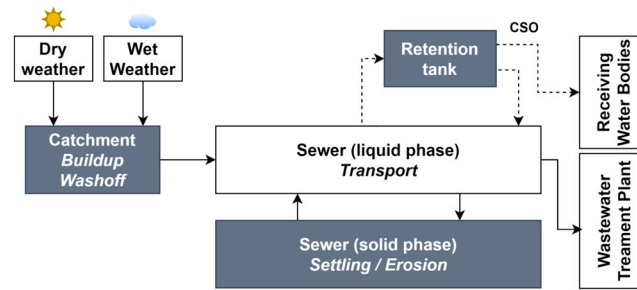


Fig. 2. General characterisation of the sewage system [79].

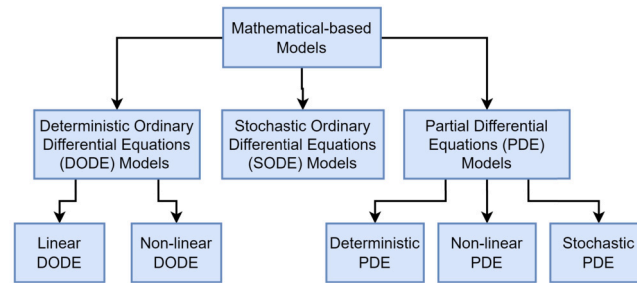


Fig. 3. Mathematical-based model Taxonomy by [98].

The composition of the sewage network includes complex features, such as subsystems with nonlinear dynamics with delays such as open-flow channels, pressured pipes, weirs, siphons and flow dividers. Moreover, the control of sewage networks is achieved through a combination of continuous variables (flows and volumes), and discrete variables (on-off control devices and overflow events) [54]. In general, the available storage volume in retention tanks is the main controlled variable in the network, which is manipulated by an actuator (valve or pump) located at the outlet of each retention tank that controls the outflow. Normally, sensors are included for level and outflow measurements, but estimators are often observed in their absence [6].

Sewage has both a liquid and a solid phase and they can be described using parameters such as Total Suspended Solids (TSS), Biochemical Oxygen Demand (BOD), Chemical Oxygen Demand (COD), Nitrates and Nitrites (NOX), Total Kjeldahl Nitrogen (TKN), Total Phosphorus (TP) and Dissolved Oxygen (DO) [62] as well as concentration of specific compounds such as ammonia (NH_4^+). The chemical and biological process of the sewage system are often viewed as a reactor where all phases interact [84] and solid particles are subjected to dynamic behaviours as described in [71] and [64].

Mathematical descriptions of pollutants and their known behaviour are important to understand these phenomena and improve sewage systems models.

2.1. Modelling taxonomy

Several papers developed strategies for modelling of sewage systems. In [98], a comprehensive review of mathematical descriptions of pollutants is presented, which summarises the mathematical-based models as in Fig. 3.

The author details the mathematical modelling methods:

1. Deterministic ordinary differential equations models (DODE)
2. Stochastic ordinary differential equations (SODE)
3. Partial differential equations (PDE)

These methods are used to describe the space-time evolution of pollutants dynamically, and the author presents several approaches, from a simple Linear DODE (LDODE), to the highly complex Stochastic PDE (SPDE). The final remark of the author expresses the need of high computational requirements for the complex methods proposed.

A more general classification was proposed in [32] based on a literature review, as seen in Fig. 4. This taxonomy is more relevant for this review paper, because it focuses on modelling techniques applied specifically for RTC design and application, considering the time constraints, computational requirements and accepting that RTC systems do not necessarily require highly complex descriptions of the infrastructure. Therefore complex models such as DODE, SODE and PDE are out of the scope of this paper. For further information in these models, the reader is encouraged to read [3], [49], [56], [71] and [98].

The taxonomy in Fig. 4 separates models based on its purpose: simulation-oriented or control-oriented. In this paper the focus is on water quality models needed for control implementation and the analysis of simulation-oriented models is disregarded.

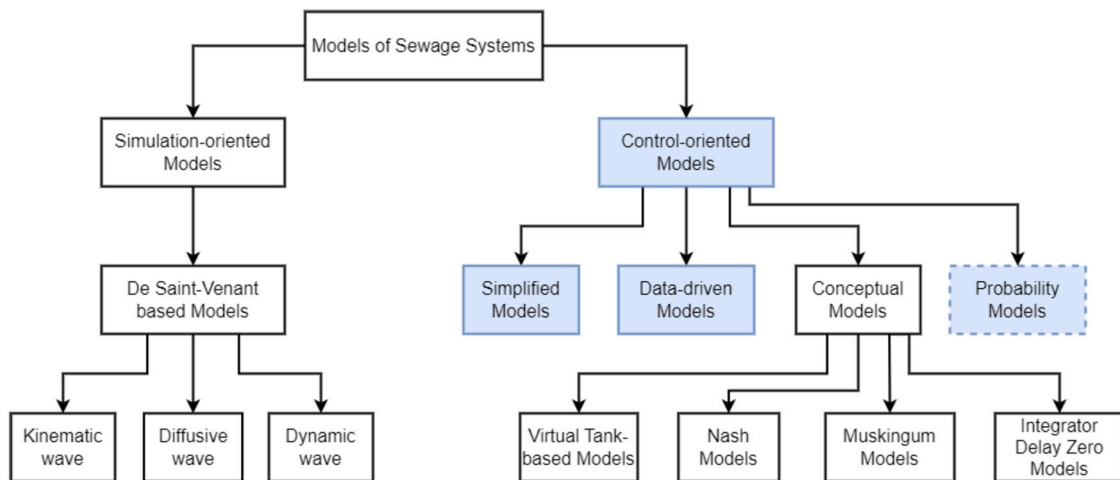


Fig. 4. Sewage system model taxonomy proposed by [32].

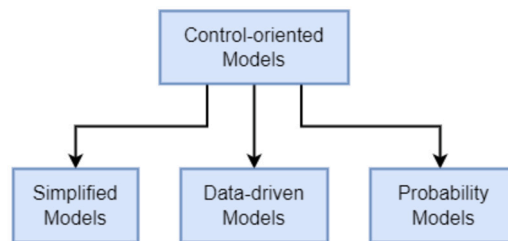


Fig. 5. Proposed sewage system control-oriented model taxonomy.

Control-oriented models are specifically designed to be used with control systems, which also needs to factor-in the relationship between accuracy as well as the type of controller applied. They are separated in three main classes:

- **Simplified models** are mathematical formulations of the sewage network with simplified dynamics for the sake of a reduced computational burden. These models are characterised by a set of equations that provide an approximate behaviour of the real system.
- **Data-driven models** use process data and machine learning techniques that provide the output of the system without any knowledge of the behaviour of the sewage network. Data-driven models are commonly known as “black-box” models, since the actual mathematical model is unknown. The main use of data-driven models for control is, therefore, prediction, especially in methods where future predictions are essential for good performance.
- **Conceptual models** describe the system through a parametrisation of the tanks in the network. Each sub-category shown in Fig. 4 represents a different way the tank is parameterised. The conceptual models are commonly used for volume-based control algorithms. Since no related conceptual model was found in the literature this type of model is not further reviewed in this paper.

Literature research revealed methods for defining a quality model based on probability [77] that are not contemplated in the current taxonomy and, therefore, a new branch is proposed in this paper as shown in Fig. 5.

- **Probability models** consider that not every variable of the process is deterministic and a solution is to describe the nondeterministic variable with a statistical representation [68].

The literature in pollution-based control showed that current modelling approaches fit the taxonomy proposed. There are several important pollution-based modelling topics, such as sewer corrosion [39] [94] and greenhouse gas emission models [27], however this paper focuses on hydrological models and details different examples of models for each taxonomy.

2.2. Simplified models

The complexity of modelling a system using phenomenological non-linear equations might prevent the model to be used for controlling the network, due to high computational cost to calculate the model’s output in real time or difficult calibration of the variables to generate a reliable model. In addition, a complex model in general also turns the controller design into a much more

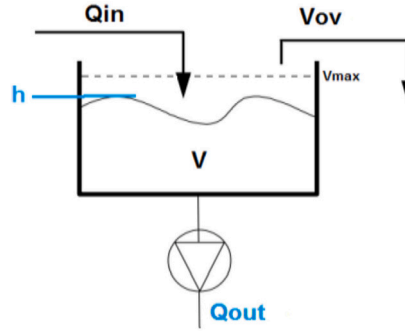


Fig. 6. Tank model for wastewater quantity modelling [48].

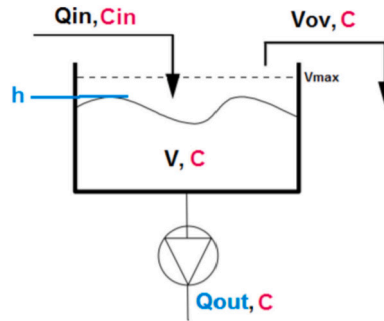


Fig. 7. Tank model for wastewater quality modelling [48].

complicated task. A solution is an approximation, which can be achieved by linearisation of the non-linear model or by assuming a simplified dynamic behaviour.

In the sewage network, an example of a simplified model is to approximate hydraulic elements, such as valves, pumps, pipes and tanks with simple functions. The dynamics of pumps and valves are commonly simplified due to their fast action compared to the slower sewage network overall dynamic. Pipes and tanks can also be simplified, for example, using a plug-flow method which considers the pollutant flow with constant velocity, volume and concentration through partially filled pipes, defined by a simple time-delay. Similarly, a tank is normally considered as a fully-mixed reactor.

In the next sections simplified models used for pollution-based control and their application are reviewed.

2.2.1. Concentration balance model

[48] defines a quality model for the sewage system, where a simple tank-based model is used with three assumptions for the pollution within the description: (i) one global indicator for pollution, (ii) a simple dilution effect and (iii) homogeneous concentrations in the tank.

Mahmoodian describes the sewage system in terms of a simple tank-based quantity model where the outflow of the tank is controlled. Each tank is modelled individually and contains a sensor for measuring the level, as depicted in Fig. 6.

The inflow Q_{in} , at time t , can be derived either by measurements or using a mass balance equation in a tank by knowing the outflow in the tank, Q_{out} , the volume V calculated using the level measurement and the volume overflow V_{ov} , as seen in Equation (1):

$$V(t) = V(t - \Delta t) + Q_{in}(t)\Delta t - Q_{out}(t)\Delta t - V_{ov}(t) \quad (1)$$

Then, by connecting each tank using plug flow, the network is fully described. This mathematical formulation of the tank model can be extended to consider the concentration of the wastewater entering and leaving the tank, as presented in Fig. 7.

Similarly to volume, the concentration C is calculated using the mass balance:

$$m(t) = m(t - \Delta t) + mf_{in}(t)\Delta t - mf_{out}(t)\Delta t - m_{ov}(t) \quad (2)$$

where m is the mass and mf is the mass flow. By considering $m(t) = C(t)V(t)$, where C is the concentration in the tank and C_{in} is the concentration of the inflow, and $mf(t) = C(t)Q(t)$, Equation (2) can be written as Equation (3):

$$C(t)V(t) = C(t - \Delta t)V(t - \Delta t) + C_{in}(t)Q_{in}(t)\Delta t - C(t - \Delta t)Q_{out}(t)\Delta t - C(t - \Delta t)V_{ov}(t) \quad (3)$$

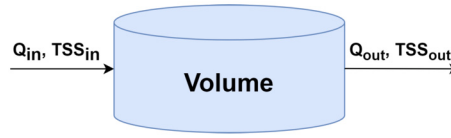


Fig. 8. Water tank container [79].

By manipulating Equations (1) and (3), and isolating $C(t)$, the concentration is modelled as Equation (4):

$$C(t) = \frac{C(t - \Delta t)V(t - \Delta t) + C_{in}(t)Q_{in}(t)\Delta t}{V(t - \Delta t) + [Q_{in} - Q_{out}]\Delta t - V_{ov}(t)} - \frac{C(t - \Delta t)Q_{out}(t)\Delta t - C(t - \Delta t)V_{ov}(t)}{V(t - \Delta t) + [Q_{in}(t) - Q_{out}(t)]\Delta t - V_{ov}(t)} \quad (4)$$

The author mentions that this model does not rely on measurements of concentration, which is beneficial for applications since it is not required to change the real network for the purpose of a pollution-based RTC [48]. However, it is necessary to estimate the concentration and, therefore, Mahmoodian proposes the use of a first order Taylor series to estimate C_{in} and calculate the uncertainty propagation from the derived approximation. The Taylor series uncertainty propagation is characterised by a U_X , which is the standard deviation of X such that variance of X is $Var(X) = U_X^2$. Therefore, the standard deviation for the concentration is given by Equation (5):

$$Var(C(t)) = U_{C(t)}^2 = \sum_{i=1}^6 \left(\frac{\partial C(t)}{\partial A_i} \right)^2 U_{A_i}^2 \quad (5)$$

where A_i indicates each variable of the system.

$$A_1 = C(t - \Delta t), A_2 = V(t - \Delta t), A_3 = Q_{in}(t),$$

$$A_4 = C_{in}(t - \Delta t), A_5 = Q_{out}(t), A_6 = V_{ov}(t - \Delta t)$$

Hence, to avoid the need for measurements, the concentration can be approximated using Taylor series and the uncertainty model from Equation (5) results in Equation (6):

$$U_{C(t)}^2 = U_{C(t-\Delta t)}^2 (X(t - \Delta t))^2 + U_{C_{in}(t-\Delta t)}^2 (1 - X(t - \Delta t))^2 \quad (6)$$

where $X(t - \Delta t)$ is defined in Equation (7)

$$\begin{aligned} X(t - \Delta t) &= \frac{V(t - \Delta t) - Q_{out}(t)\Delta t - V_{ov}(t)}{V(t)} \\ &= \frac{V(t) - Q_{in}\Delta t}{V(t)} \\ &= 1 - \frac{Q_{in}(t)\Delta t}{V(t)} \end{aligned} \quad (7)$$

Moreover, considering $Q_{in}(t) = Q_{DW}(t) + Q_{WW}(t)$, of which Q_{DW} is the inflow derived from dry weather and Q_{WW} is the inflow from wet weather, resulting in the uncertainty model in Equation (8):

$$U_{C_{in}(t)}^2 = U_{C_{DW}(t)}^2 \left(\frac{Q_{DW}(t)}{Q_{in}(t)} \right)^2 + U_{C_{WW}(t)}^2 \left(\frac{Q_{WW}(t)}{Q_{in}(t)} \right)^2 \quad (8)$$

Using Equations (6) and (8) it is possible to calculate the uncertainty of the unmeasured concentration in the tank and use this information for controlling the sewage system with a pollution-based RTC.

This modelling is relatively straightforward, easily incorporated in an optimisation algorithm, however the assumptions of a global concentration indicator overly simplify the model in that it makes no distinction between soluble compounds and particulate matter.

2.2.2. Total suspended solids balance model

Instead of modelling the concentration using a global indicator, some authors propose the use of specific analytes, such as TSS or COD. This is the case for the model in [79], which initially defines a hydraulic model assuming a water tank container as in Fig. 8.

Sun proposes three different modelling equations for the TSS:

$$\textbf{Model 1: } TSS_{out}(t + 1) = (1 - c)TSS_{out}(t) + cTSS_{in}(t) \quad (9)$$

where TSS_{in} and TSS_{out} are the TSS entering the structure and the TSS leaving the structure, respectively. Equation (9) can be generalised such that the calibration parameter c can be independent for TSS_{in} and TSS_{out} , showed in Equation (10):

$$\textbf{Model 2: } TSS_{out}(t + 1) = c_1 TSS_{out}(t) + c_2 TSS_{in}(t) \quad (10)$$

which c_1 and c_2 are calibration parameters for TSS_{out} and TSS_{in} respectively. Furthermore, as mentioned previously, the TSS is affected by sedimentation, erosion and time delays in its dynamics. Thus, Sun upgraded the model to add this information and improve the characterisation, resulting in Equation (11a) and (11b):

$$\text{Model 3: } TSS_{out}(t) = c_{vc} TSS_{in}(t - d) + e_p \quad (11a)$$

$$Q_{out}(t) = t_{vc} Q_{in}(t - d') \quad (11b)$$

where d is the delay of TSS, d' is the delay of the flow and c_{vc} , t_{vc} and e_p are parameters that need calibration. Sun et al. discusses that, using these modelling methodologies, it is possible to develop a pollution-based RTC, although it is necessary to have an online calibration of several parameters, which are done via a general algebraic modelling system (GAMS) using historic or real-time data from telemetry systems. In a comparison between the three proposed models, all three behaved similarly, whereas Model 2 was slightly better. It is suggested to use a simulation software such as SWMM [69] for the real-time calibration of parameters. For that reason, these models are highly sensitive to changes in the sewer system and need to be actively calibrated to guarantee a reliable prediction.

2.2.3. Effluent quality index model

The last simplified model presented establishes an effluent quality index that is used to measure the pollution load, integrating the following pollutants together: TSS, BOD, COD, NO_x , TKN and TP. The model characterisation can be seen in [42] and is further described in [62]. It is a unique approach by modelling several types of pollutants simultaneously, allowing a more complex control of the sewage system. The effluent quality index (EQI) is defined by Equation (12):

$$EQI = \frac{1}{1000(t_f - t_0)} \int_{t_0}^{t_f} C_t Q_e(t) dt \quad (12)$$

where $Q_e(t)$ is the overflow rate, t_f and t_0 are the final and initial instants, respectively, and:

$$C_t = \beta_{TSS} C_{TSS} + \beta_{COD} C_{COD} + \beta_{BOD} C_{BOD} + \beta_{NOX} C_{NOX} + \beta_{TKN} C_{TKN} + \beta_{TP} C_{TP}$$

where C_x and β_x corresponds to the concentration and weights for each pollutant x , respectively.

The quality index modelling increases the complexity of the optimisation by increasing the number of variables, while also requiring measurement (or estimation) of several concentration parameters. This introduces more uncertainties to the optimisation where those measurements are unreliable or nonexistent, degrading the control of the network. However, this technique greatly improves the versatility of the algorithm, allowing to control different types of pollutants independently. The model also enables the control algorithm to be customised depending on the requirements of the network, increasing the importance of certain pollutants for the controller.

2.2.4. Section summary

Simplified models are generally used in sewage system applications which have highly non-linear and complex dynamics, approximating the system with sufficient accuracy for control. Although the assumptions needed for a simplified model might seem excessively simplistic, for real time control, these models are attractive because of their low computational effort and easy implementation.

2.3. Data-driven models

The use of data and machine learning techniques is the foundation of data-driven models. When using simplified models, information of the behaviour of the system is lost for the sake of faster computation and, also, the calibration of model parameters is hard to calculate due to uncertainties [4]. Assuming that the model is perfectly calibrated, there is no guarantee that the calibration remains unchanged over time. In fact the system is constantly changing due to shifts in wastewater patterns, seasonal runoff dependencies and land use changes [86]. An alternative is to use data-driven modelling approaches, which rely on real time data processing to represent the dynamics of different variables of the wastewater network, either by using data from software simulation, on field measurements or historical data [83].

The continuous evolution of sensors [8,14] has led to an increasing number of studies and implementations of data-driven models, encouraging the application in real-time [22]. A few papers can be found that summarise data-driven models for pollution-based systems; [58] reviews detection methods and indicator parameters for quality measurements with an extensive detailing of each type of pollutant's measuring device and a short description of each; [37] reviews and proposes source identification methods for tracking illicit discharges with low cost, showing a good precision for cost-effective analytical techniques; [38] details the modelling of water quality in sewers, providing an extensive review for sewage systems including data-driven models of different purposes.

Data-driven models are easier to adjust to different types of water quality data than other models. The main reason for that is the ability to derive a suitable model regardless of the behaviour of the system through the use of process data and algorithms that mimic human learning. In this section, data-driven models utilised in combination with RTC systems are presented. One notable work is the publication of [1], which studies different deep learning algorithms to predict several water quality parameters. The paper proposed

Table 1
Variables and their definitions [1].

Variable	Definition
d_s	Avg. sediment depth in a sewer (cm)
NOC	Number of connections to a sewer
D	Diameter of the sewer
Q	Sewer capacity
S	Sewer gradient (%)
q_{peak}	Peak flow (m^3/s)
L	Length of the sewer (m)
V_{peak}	Velocity that occur at peak flow (m/s)
$L_{cum.}$	Cumulative sewer length (m)
d_{peak}	Depth of flow at peak flow (m)
$L_{dir.}$	Cumulative sewer length (m)
τ_{peak}	Average tractive stress at peak flow (pa)

several models used to predict sedimentation and compare the results using five deep learning algorithms: Multi Linear Regression (MLR), Multilayer Perceptron (MLP), Recurrent Neural Network (RNN), Long Short-term Memory (LSTM) and Recurrent Gate Unit (GRU).

2.3.1. Multiple linear regression and neural network techniques for sediment prediction

[1] studied and compared two different data-driven models, a MLR based on [2] and a neural network model based on [21], where both methods are used for predicting in-sewer location and the sediment deposition. It presents an assessment of these models to find the sediment accumulation, where the models inputs and outputs are shown in Table 1, given that the sediment depth, d_s , is the dependant variable.

The methodology for both methods is similar and is generalised as shown in Fig. 9. First, the data collection and processing is relevant to determine the variables of the process and perform preliminary analysis of the data. This processing will help to improve the quality of the data and facilitate the use of the data set. The next step is the design of the model, that comprises the selection of several methods that will use the data and create models. In sequence, the estimation and assessment of each model is done and a data set is generated specifically for testing. Finally, a model is selected and validated, defining the most efficient model and using different scenarios for the validation.

2.3.2. Multiple linear regression model

The Linear Regression method is a technique where a single output variable is derived using a simple linear equation from a single input. A more generalised formulation, called Multiple Linear Regression (MLR), approximates the output by multiple inputs simultaneously. The output of the regression model, Y , is calculated by using the following model for n points collected [43]:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \beta_p X_{ip} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (13)$$

where β_0 is an offset variable, β_1 , β_2 and β_p are the regression coefficients for the independent variables X_1 , X_2 and X_p , respectively and ε is the error associated with the model. In order to calculate the coefficients X , the model in Equation (13) is rewritten to be represented in matrix form, which yields Equation (14):

$$Y = X\beta + \varepsilon \quad (14)$$

Herein, the output Y and error ε are n -dimensional vectors, the input matrix X is $n \times (p + 1)$ and the coefficient vector β is $(p + 1)$ -dimensional. Using a least-square regression, the coefficient β can be found by Equation (15):

$$\beta = (X^T X)^{-1} X^T \quad (15)$$

Al-Ani implemented the algorithm for a specific process in Baghdad where a sediment survey was carried out in the sewer system of the city. The implementation of the method was done using the software SPSS Neural Network [15] developed by IBM, which resulted in Equation (16), a linear model using the MLR algorithm as below:

$$d_s = 0.036L_{cum.} + 0.203L + 6.685NOC - 1.142Z - 0.022L_{dir.} - 48.493 \quad (16)$$

By using collected data, the MLR model is able to predict sediment accumulation throughout the sewer, using a simple approach with very low computational cost. An advantage of this method is that little knowledge of the behaviour of the sewer is required and as long as new data-sets are acquired, the models coefficients are easily re-identified to improve the performance, since it is known that the dynamics of a sewage network change over time. However, the main issue with this approach is the reliability of the data used for the identification and also the error inherent to measuring systems, which increases the error of the MLR method if the sensors used in the data collection are not correctly placed or calibrated.

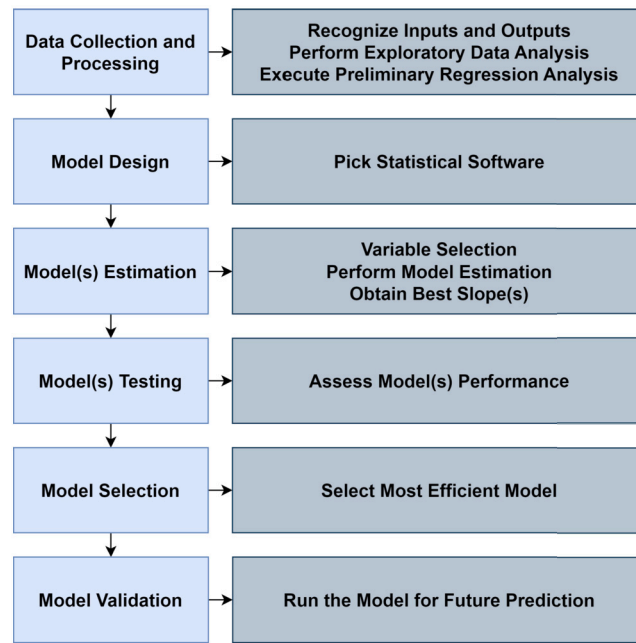


Fig. 9. Methodology for MLR and ANN methods [1].

2.3.3. Artificial neural network model

Modelling sewage systems with Artificial Neural Network (ANN) is a method often seen in the literature. It is mainly applied for regression and classification [36]. The ANN emulates the human brain using underlying relationships between data sets, using an architecture comprised by a Input Layer, a number of Hidden Layers and an Output Layer [1]. The most common architecture is the Multilayer Perceptron (MLP), a feedforward architecture which uses backpropagation as the learning algorithm.

The data used in the MLP approach is divided into three sets: a training set which is used to find the weights of the neural network, the testing set, responsible for checking the error and avoid over-training of the network and the validation set that is used for validation of the model derived from previous training.

The MLP-ANN proposed by Al-Ani et al. was identified using the tool SPSS Neural Network and the resulting model for the same sewer in Baghdad resulted in one neuron with a single hidden layer, while the weights of the network were optimised using a gradient descent method as criteria for training the neural network model. Equation (17) is the final result in a mathematical expression:

$$d_s = 1.716 * \tanh(X) + 0.04 \quad (17)$$

where X is defined in Equation (18) below:

$$\begin{aligned} X = & 0.024 + 0.114NOC + 0.027L - 0.262S \\ & + 0.758D - 1.495q_{peak} + 0.207V_{peak} \\ & + 0.326Q + 0.304q_{peak}/Q \\ & + 0.325d_{peak}/D + 0.148\tau_{peak} \end{aligned} \quad (18)$$

There are several advantages of using a ANN-MLP, but, specifically for a sediment accumulation model, the ability to adapt to complex relationships between dependant and independent variables and to changes that occur in the process is very unique [11]. Besides that, there are many open-source software packages available for implementing ANN's, such as scikit-learn, a simple library of machine learning methods using python [59] and OpenNN, a neural network library for C++ [50]. The main disadvantage of using neural networks is finding the best network structure, i.e. the selection of number of layers and number of neurons in each layer, as well as the selection of other parameters of the neural network, which in most cases is done by trial and error although there are optimisation techniques that solve this issue. Another disadvantage is the lack of intuition on the model, since it is difficult to understand the reasoning behind the solution provided by the network.

2.3.4. Deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks

A similar solution developed in the data-driven modelling framework was proposed by [40], which also used a similar approach to the MLP proposed by Al-Ani, however with the objective of using multi-source data fusion and deep machine learning models to predict several pollutants, such as BOD, COD, NH_4^+ , TN, and TP.

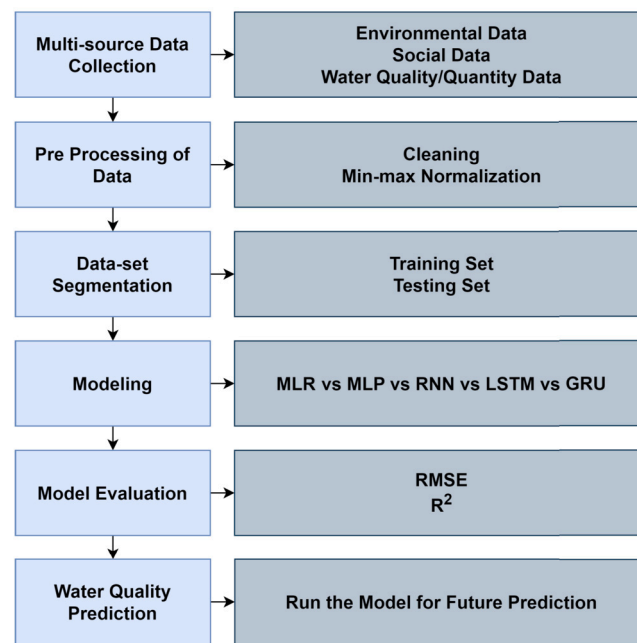


Fig. 10. Framework for water quality prediction [38].

Jiang defines the framework as shown in Fig. 10. The first step starts with data collection from several sources, such as (1) environmental data, (2) social data, (3) water quality data and (4) water quantity data. Then, the multi-source data is pre-processed and separated in a training and testing set. Finally, the author compares several deep learning algorithm to predict the water quality. The algorithms the author studied are MLR, ANN-MLP, Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU).

The case study was realised using data from a southern city in China. The input dataset consists of: land area, drain pipe diameter, population, drinking water supply, flow, volume, level, pH, temperature, conductivity and suspended solids. The output indicators are BOD, COD, NH_4^+ , TN, and TP.

The comparison of the models was done using Root Mean Squared Error (RMSE) and R^2 metrics. The conclusions are the following:

1. MLP is more suitable for water quality prediction than MLR because traditional machine learning algorithms are better for multi-source data than linear methods.
2. RNN-based methods are better than MLP because those methods have an advantage when the data is time series data.
3. Among the deep learning methods, GRU showed a significant better performance than RNN and LSTM for most indicators.

The author concludes that, although the prediction of water quality using deep learning algorithms has great potential, these models depend on the input indicator selected from the data set used and it is crucial to understand which of these indicators are key for training-learning models.

2.3.5. Section summary

The exponential improvements in computer performance and artificial intelligence increases potential practical implementation of deep learning models with big data, therefore the research on water quality prediction with deep learning has a potential to be key for the future. The ability to adapt to complex dynamics without complete knowledge of such phenomena is appealing but has its limitations.

2.4. Probability models

The use of simplified and data-driven models introduces errors due to uncertainties, which means that the prediction using these types of model for controlling purposes might be less reliable. An alternative modelling technique is using models that factor these uncertainties using probabilistic assumptions of the system, also known as stochastic models [68]. The difference between the deterministic and stochastic model is that the later considers the description of the uncertainty within its formulation using a statistical approach.

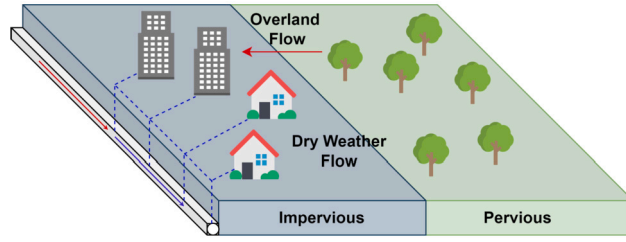


Fig. 11. Representation of a catchment by [52].

2.4.1. DO and BOD transport model

Water quality indicators such as DO and BOD have a big impact on receiving bodies of water when there is wastewater overflow. Their transport in sewer systems is a complex process and is usually approached in modelling systems by assuming a fixed concentration or by estimation using isolated measurements [90]. However, that leads to the effect that information regarding the concentration during storm events are lost. The solution proposed by [52] is to use a probabilistic model that is able to describe the dynamics of the process. This reduces uncertainties because it does not require a complex parametrisation of the complete sewer network and the model only needs an input value for each catchment. The catchment is shown in Fig. 11, where the overall flow is the combination of a dry weather flow and an overland flow, which is separated as a pervious and impervious overland flow.

Firstly, Morales defines the flow path using state variables, i.e., the pervious overland state ($x_{oi,perv}$), impervious overland state ($x_{oi,imp}$) and the dry weather flow state (x_{DWF_i}), as well as a series of conduit states (x_c), where $i = 1, 2, \dots, n$ and n is the number of catchments. Then, in Equation (19) is shown the probability of the load in each flow:

$$\begin{aligned} P_{x_{oi,perv}}(t) &= \frac{\xi_{i,perv}(t)}{\xi_T(t)} \\ P_{x_{oi,imp}}(t) &= \frac{\xi_{i,imp}(t)}{\xi_T(t)} \\ P_{x_{DWF_i}}(t) &= \frac{\xi_{DWF_i}(t)}{\xi_T(t)} \end{aligned} \quad (19)$$

where $\xi_{i,perv}$, $\xi_{i,imp}$ and ξ_{DWF_i} are the total constituent loads from the i th pervious, impervious and dry weather flow regions at time t , respectively.

Morales then describe a travel time probability distribution $f(t')$, where t' is a random variable associated with the travel time of a drop of water in the network, and where $f(t')$ allows to derive the concentration in the outlet, by propagating the concentration through the network. The calculation of $f(t')$ extends the work of [70] by taking into account non-conservation of mass in the BOD and DO during the travel time. Finally, the combined pollutograph at the outlet of the sewer network is calculated by convolving the travel time probability distribution and the total load, given by Equation (20):

$$\xi(t) = \sum_{t'=1}^{\infty} [f(t, t') \times \xi_T(t)] \quad (20)$$

The model proposed proved to be an alternative formulation for prediction of BOD and DO in control systems. When comparing with the benchmark software SWMM, the performance of the model showed a good accuracy while requiring less information and time for the implementation. However, this method is much less intuitive and the use of probabilistic variables is not a straightforward approach as the estimation of such variables are complex and harder to grasp.

2.4.2. Bayesian multi-objective calibration for TSS predictions

The representation proposed by [76] is a stochastic model for TSS prediction where the goal is to improve the model errors by using statistical description of the bias. It uses an auto-regressive error model to provide more reliable estimation of TSS and reduce the uncertainty of TSS using a Bayesian multi-objective approach to improve calibration of the model. The general mathematical description of a stochastic model is given by Equation (21):

$$\tilde{Y}(X, \theta^M, \theta^e) = Y_M(X, \theta^M) + \tilde{B}_M(X, \theta^e) + \tilde{E}(\theta^e) \quad (21)$$

where $Y_M(X, \theta^M)$ is the deterministic output of the model, $\tilde{Y}(X, \theta^M, \theta^e)$ is the real output of the model, i.e., the sum of the deterministic model with the error, hence $\tilde{Y}(X, \theta^M, \theta^e)$ is a random variable. The inputs of the system are X , the parameter of the deterministic model is represented by θ^M and the parameter of the error is θ^e . The variable $\tilde{B}_M(X, \theta^e)$ is used to describe the model bias and $\tilde{E}(\theta^e)$ describe the random measurement noise.

The output \tilde{Y} is modelled as a random variable and the author defined the prediction as a predictive probability distribution $p(\tilde{Y}|X)$, which is calculated using a Bayesian approach:

$$p(\tilde{Y}|X) = \int p(Y^0|\theta, X)p(\theta)d\theta \quad (22)$$

where $p(\theta)$ represents the prior knowledge of the parameters θ , $p(Y^0|\theta, X)$ is the likelihood function of the model that measures the probability an output Y^0 is generated by the model given an input X and the parameter chosen randomly from $p(\theta)$.

Using Equation (22), the model proposed by Sikorska to predict the TSS can be derived, resulting in Equation (23):

$$\tilde{C}_{TSS}(P^0, \theta) = C_{TSS}(P^0, \theta) + \tilde{B}_{M_{BW} \times M_{RR}}(P^0, \theta) + \tilde{E}_{TSS}(\theta) \quad (23)$$

where $\tilde{B}_{M_{BW} \times M_{RR}}$ comprises the systematic prediction errors of the concentration of TSS, C_{TSS} , due to build-up/wash-off (BW) and rainfall/runoff (RR) model components and \tilde{E}_{TSS} describes the random errors of the prediction of TSS. Likelihood function of the TSS model is given by $p(C_{TSS}^0|\theta, P_0)$.

The author further discusses the model by including multiple outputs, such as flows and precipitation. This modelling approach showed an improvement in the prediction efficiency of the TSS model and the bias description led to a more reliable uncertainty estimation. The main challenge of the approach is formulating the model bias using prior information, since the bias is defined to compensate different error sources, which not always have a direct physical interpretation.

2.4.3. Section summary

Probability models for quality prediction are an interesting approach to consider uncertainty in the formulation of the control. Such formulation can increase robustness and improve the performance of the controller in scenarios where randomness is known with good accuracy.

3. Pollution-based real time control

Generally, the main goal of Pollution-based Real Time Controllers (P-RTC) is to reduce environmental impact of the wastewater network by introducing pollution information in the controller. The difference between different P-RTC approaches is how this information is handled within the formulation of the controller, the type of pollutant controlled and the framework used to describe the pollution. In this paper is proposed the following P-RTC methods classification:

- **Rule-based Real Time Control:** the control algorithm is defined using heuristics. These heuristics are usually in the form of rules that are imposed to the problem and expect a specific behaviour if the rules are followed. A pollution-based approach using this formulation defines several rules applied to the sewage control system, for example, the algorithm checks if the concentration in a given basin is higher than a certain value and proposes an action depending on the output of this rule (which is generally a 'yes' or 'no' output).
- **Optimisation-based Real Time Control:** in this type of method, the overall objective is achieved by minimising (or maximising) a function using optimisation methods, which will provide an optimum operation of the system. Each method can have different objective functions depending on the desired behaviour of the system. One example is a P-RTC based on Model Predictive Control (MPC) which is an algorithm that optimises a cost function in order to control the system over a prediction horizon by using a model to predict the behaviour of the system. In the case of a pollution-based scenario, a common objective defined in the optimisation is to minimise the concentration during overflows of tank structures.
- **Reinforcement Learning-based Real Time Control:** through a reinforcement learning (RL) method, an algorithm learns the behaviour of the system by a trial-and-error experience and by rewarding control actions that meet specific goals. This reward based structure will steer the algorithm towards the actions that maximise the output. For example, A RL-based RTC can be developed to improve the water quality by rewarding control actions where the concentration of overflowed wastewater is reduced [10].

3.1. Rule-based RTC

Due to high complexity of the dynamics of sewage systems, the definition of a model is a difficult task to be achieved. Therefore, efforts have been done to create real time control systems which avoided the use of models and one example are rule-based systems, where the control strategy is based on heuristics [96]. Control algorithms based on heuristics are usually more intuitive for operators since the control rules are based on logical statements using expert knowledge of the behaviour of the sewage system. Also, rule-based RTC systems do not require the definition of an optimisation problem and, therefore, the control is less complex and demand less computational power for computing the solution. Combining all these advantages it is clear that rule-based RTC is an attractive solution for controlling sewage systems and several researches focused their effort specifically in designing rule-based RTC systems in an pollution-based framework [96].

3.1.1. Water quality data-informed real-time control

The algorithm proposed by [72] adds a real time controller to the normal operation of the sewage system by adding rules to the behaviour of the system and using the available data. The author defines this controller as a water quality-informed real-time active control due to the fact that the algorithm uses quality (and rain) monitoring to derive a RTC based on rules that are applied and are actively controlling the sewage system.

The author defines four possible control rules in his study, a passive control, a retention control, an on/off control and a TSS control, as can be seen in Table 2. Each of these rules have a different impact on the behaviour depending on the current state of

Table 2
Control rules implemented in [72].

Type	Description
Passive Control	Valve always open
retention Control	If an event occurs, valve opening = 0% Else, valve opening = 100%
On/off Control	if $h < h_c$, valve opening = 0% if $h \geq h_c$, valve opening = 100%
TSS Control	if $C \geq C_c$, valve opening = 0% if $C < C_c$, valve opening = 100%

the system and the different variables of the sewage system. The baseline operation is the passive control that is an operating mode where the actuator (a valve) is constantly open.

The retention control mode differs by monitoring inflow events and acting whenever a event occurs, which activates the retention controller and completely closes the valve. The valve is kept closed for a specific duration t_d , called the retention time, with the objective of storing the inflow volume into the structure. The third mode, on-off control, defines a threshold value for the level of the tank structure, called h_c , and if the water level is below h_c , the valve is remained closed, however when it reaches the threshold, is completely opened to drain the volume within the tank.

These two modes are traditional volume-based algorithms, thus, to improve the quality of the sewage, Sharior proposes a fourth mode that takes into account the concentration of TSS inside the tank, called the TSS Control mode. The general idea is similar to on-off control mode, the difference is the threshold refers to the TSS concentration, called C_c , and the logic of the on-off actuation is different. In the TSS control, when the concentration of TSS is above a threshold concentration, the controller is fully closed, otherwise, the valve is opened. The goal of this procedure is to increase dilution when rain events occur to decrease the concentration and when the concentration reaches a specified value, the valve is opened and drains the tank, sending this volume towards the treatment plant.

The method proposed does not use an explicit model, the author clarifies that measurements were used for the formulation, implementation and simulation of the algorithm.

3.1.2. Quality control based on retention time

An interesting approach for controlling a urban drainage system using a rule-based algorithm was presented by [74], differing from the previous algorithm by using only a volume-based set of rules to influence the quality of the wastewater within the drainage system during dry-weather periods. The goal of this algorithm is to control the settling process in the stormwater basin and, as a result, influence the TSS retention time using a generic formulation that considers a required volume in the basin and information regarding runoff and rainfall [74].

This algorithm is activated during dry-weather periods, whereas during wet-weather, a volume-based optimisation is implemented without any quality aspects. Therefore, since this review paper is only interested in the RTC systems that are directly impacting the pollution in the receiving water bodies, here, the only aspect discussed is the implementation of the, dry-weather, rule-based controller. The author defines four rules to control the wastewater system, as showed in Equation (24) below:

$$\begin{aligned}
 &\text{if } t_{\text{nexttrain}} \leq t_e : \\
 &\quad \rightarrow Q_t = Q_{\text{max}} \\
 &\text{if } t_e < t_{\text{nexttrain}} \leq t_e + 20h : \\
 &\quad \rightarrow Q_t = Q_{\text{max}} * \frac{t_e}{t_{\text{nexttrain}} - t_f} \\
 &\text{if } t_e + 20h \leq t_{\text{nexttrain}} < 40h + t_e^{\text{max}} : \\
 &\quad \rightarrow Q_t = Q_{\text{max}} * \frac{t_e + 20h}{t_{\text{nexttrain}} - t_f} \\
 &\text{if } t_{\text{nexttrain}} \geq 40h + t_e^{\text{max}} : \\
 &\quad \rightarrow \begin{cases} Q_t = 0 & \forall t < 40h \\ Q_t = Q_{\text{max}} * \frac{t_e}{t_e^{\text{max}}} & \forall t \in (40h, 40h + t_e^{\text{max}}) \end{cases}
 \end{aligned} \tag{24}$$

with t_e being the emptying time of the basin until availability of the storage volume V_{req} at maximum outflow $Q_{\text{max}}(s)$. The author assumes that the main characteristics of the sewage systems are known, such as the inflow to the basin, and therefore no model for control is used.

An important advantage of using this method, as mentioned by the author, is that the use of this set of rules ensures that the available volume within the basin is sufficient to handle the predicted upcoming rain event without overflow, which reduces the discharge of highly concentrated volumes. However, the main disadvantage with this approach is that the control is considered

Table 3

Papers on rule-based RTC methods with pollution aspects.

Authors	Year	Type of Control	Description	Type of Pollutant
[93]	1997	Local Pollution Rule-based + Global Volume-based Optimisation	Minimisation of overflow concentration and tracking of pollution discharge to WWTP using a global volume-based optimisation and rule-based local concentration-based control.	Multiple pollutants
[97]	1999	Global	Fuzzy logic control and genetic algorithms are applied to achieve improved pump operations in a combined sewer. It is concluded that current pump operations can be improved by adding the water quality to the input variables.	Global indicator
[87]	2005	Global	Rule-based control of WWTP flow overload using measurement of ammonia in the river.	NH ₄
[29]	2005	Global	A rule interpreter was developed to make control decisions based on rules (if-then relations), which are evaluated with the aid of fuzzy logic and input values such as water levels, flows, etc.	COD
[35]	2011	Local	A real-time control strategy for separation of highly polluted storm water based on UV-Vis online measurements	TSS using real-time water quality measurement
[30]	2013	Local	Rule-based retention control improving water quality by maximising the retention time of water in the pond in order to increase the TSS removal efficiency through sedimentation.	TSS
[75]	2021	Local Pollution Rule-based + Global Volume-based Optimisation	Rule-based combined with a predictive dynamic control for optimal operation of stormwater system regulators during rainfall periods, while applying the heuristic control for water quality control to detain runoff in the basin during dry weather periods.	Sedimentation

only on a local scale, the downstream areas are not considered globally. Therefore, this method is unable to find a global optimum operation for the sewer system as a whole, which would greatly increase the complexity of the overall method and new rules would need to be applied, whereas the intuitiveness commonly found in rule-based algorithms is lost due to a higher number of rules necessary.

3.1.3. Double-gate storage scheme based on water quality

The next rule-based RTC described was developed by [92], called a double-gate storage scheme for water quality control. The author mentions that traditional interception-storage tank control designs are unable to deal precisely with the pollution problem in combined sewer overflows, therefore, they developed a method which directly influences the variance of a pollutant with the flow-rate. In the paper Wei presents the solution applied specifically for controlling the COD concentration, defining rules in a double-gate scheme. The COD concentration is monitored in the system's storage tank and well chamber and, using this measurement, the rule-based controller will act accordingly.

The well chamber has an intercepting weir that, when opened, directs the inflow directly to a receiving body of water. Also, the chamber is connected through a intercepting pipe to the storage tank which will store the stormwater. During dry-weather, the double-gate is offline, the intercepting weir is closed and the inflow is directed to the municipal sewage pipe. However, when rain events occur, the controller comes online and starts its operation depending on the current COD concentration measured.

- Firstly, a COD concentration threshold C_i is defined for the concentration within the well chamber, which indicates a value for the COD that is safe to discharge.
- If the concentration within the well chamber is below the C_i , this flow is directed to the receiving body water and safely discharged.
- However, when the current COD measurement is above C_i , the waste intercepting pipe is opened and the waste is stored in the tank while there is enough capacity.
- Eventually the waste in the storage tank is directed to the municipal sewage pipe.

The method developed by Wei showed a great reduction in the concentration of COD in the discharge and also a great improvement over traditional schemes commonly used for dealing with storm-water overflows. The method uses measuring devices to act directly on the concentration, which is a promising technology but which is still complex and expensive for in-situ and on-site quality assessment [19]. The tuning of this method is relatively simple, needing only the threshold COD, but similarly to the previous method, is limited to local storage control and does not provide a globally optimum solution.

3.1.4. Section summary

In Table 3 several papers are shown with more implementations of rule-based RTC. They are separated in Local methods, indicating that each CSO is controlled individually and Local Pollution Rule-based + Global Volume-based Optimisation, that differ by integrating the pollution rule-based local RTC with a volume-based optimisation. In general, rule-based methods are simple and show

improvements in local structures regarding load management. However, the capability of globally observing the network system is limited and to design a global rule-based approach requires complex interactions of rules. In the papers reviewed, water level and quality sensors are included in the formulation of the control algorithm, instead, a simplified or data-driven model could overcome situations where measurements are not available.

3.2. Optimisation-based RTC

Optimisation-based control is a method that relies on an optimisation problem which determines an optimum behaviour based on control objectives through a minimisation (or maximisation) [32]. The advantage is the possibility to control the system globally and the additional use of constraints, which restrict the optimum space solution to guarantee that the control will operate within physical requirements (e.g., a valve actuator will not open more than 100%). The difference between types of optimisation-based RTC is the method used to solve the optimum problem, which might demand a non-linear algorithm, for example, or bayesian optimisation which relies on black-box models. The objectives defined for each method may also change, depending on the desired behaviour.

The optimisation-based RTC can be generalised in state-space representation as Equation (25) [47]:

$$\begin{aligned}
 & \min_{u(n)} J(n) \\
 & \text{s.t.} \\
 & h(x(n), u(n), w(n)) \geq 0 \\
 & g(x(n), u(n), w(n)) = 0 \\
 & x_{\min} \leq x(n) \leq x_{\max} \\
 & u_{\min} \leq u(n) \leq u_{\max}
 \end{aligned} \tag{25}$$

where x , u and w are the states, controlled variables and disturbances, respectively, J is the cost function, h and g are functions that represents the inequality and equality equations, respectively, used in the constraints.

Optimisation-based P-RTC methods, the difference is key to understand the control possibilities. The main contribution is the objective function, which will determine the overall control problem and how the sewage network will behave. The objective might be to reduce overflow. Pollution-based methods are distinct for the use of pollution information, therefore, the objectives of the optimisation problem are extended or changed to deal with water quality aspects, such as minimise load overflow or maximise the load inflow to the WWTP.

3.2.1. Pollution-based MPC for TSS

Policies and regulations on water quality have been made by individual environmental agencies which increase the necessity for specific pollutants to be tighter controlled and monitored. Therefore, new strategies are being studied with the objective of developing RTC systems that are capable of controlling specific pollutants in sewers. In this context, [80] proposed a MPC strategy which seeks to minimise the total amount of pollutants released to receiving bodies of water, focusing on the total suspended solids (TSS), with the justification that this pollutant can be easily measured continuously online.

The pollution-based objective defined by Sun is given by the following function in Equation (26):

$$J = \sum_{i=n}^{n+H_p} \lambda \phi_1(i) + \beta \phi_2(i) + \alpha \phi_3(i) + \gamma \phi_4(i) \tag{26}$$

ϕ_1 , in Equation (27), is an objective used for minimising the combined sewer overflow (CSO) in the sewage system:

$$\phi_1(n) = \Delta t \sum_{i=1}^{n_{CSO}} q_{CSO}^i(n)^2 + q_{bypass}(n)^2 \tag{27}$$

where n_{CSO} is the number of CSO-diverting points, q_{CSO}^i is the overflow from the $i - th$ point and q_{bypass} is the bypassed overflow from the WWTP when the inflow to the treatment plant exceeded capacity.

ϕ_2 , in Equation (28), is an objective that tries to keep the flow incoming to the wastewater treatment plant (WWTP), q_w , close to the maximum allowed inflow to the WWTP, q_{lim}^{WWTP} :

$$\phi_2(n) = \left(q_w(n) - q_{lim}^{WWTP}(n) \right)^2 \tag{28}$$

ϕ_3 , in Equation (29), is an objective to smooth variations in the output variable by minimising the difference between the current optimum output $u(n)$ with the previous optimum $u(n-1)$:

$$\phi_3(n) = \sum_{i=1}^{n_g} (u^i(n) - u^i(n-1))^2 \tag{29}$$

where n_g is the number of controlled decision variables of the optimisation problem.

Finally, ϕ_4 , in Equation (30), is a quality based objective where the load in overflow of TSS is minimised:

$$\phi_4(n) = \sum_{i=1}^{n_{CSO}} \text{tss}_{CSO}^i(n) q_{CSO}^i(n) + \text{tss}_{bypass}^i(n) q_{bypass}^i(n) \quad (30)$$

where tss_{CSO} is the TSS concentration of the CSO and tss_{bypass} is the TSS concentration at the WWTP inlet.

The final optimisation problem is then written in a form similar to Equation (25), where the constraints are established by the physical requirements of the urban drainage network and the model used internally by MPC.

The method uses a simplified model as described in Section 2.2.2.

The method proposed by Sun et al. is simple but effective: The addition of an objective to minimise the load during overflow impacted the behaviour of the controller to improve the quality of the water discharged in receiving waters. The complexity of this method will be relative to which model is used to calculate the predicted value for the TSS (a linear or nonlinear model) and will also depend on if the monitoring of the TSS is done via measurement or estimation. The author mentioned that the method reduced the overall overflow compared to a rule-based approach. When compared to a purely volume-based MPC, the pollution-based MPC decreased the concentration of the overflowing wastewater, thus confirming that a load reduction is possible by adding a quality objective. However, it is mentioned that a certain degree of uncertainty is present and should be addressed in the optimisation to improve the control of the system. The next method reviewed suggests a solution to this uncertainty issue.

3.2.2. Pollution-based MPC considering uncertainty propagation

[48] proposed a method for a real time controller using MPC in a pollution-based framework where a global concentration indicator is added to the optimisation problem as a decision variable. Using this information, the RTC implemented is able to directly affect the behaviour of the pollution within the network in an attempt to reduce the concentration of the discharged wastewater. In order to use concentration in the formulation, the model based on volumes and flow is extended to also consider the concentration and, therefore, be able to predict it. Mahmoodian designed the pollution-based MPC as in Equation (31):

$$\min J = \min \sum_{i=n}^{n+H_p} \lambda \phi_1(i) + \beta \phi_2(i) + \alpha \phi_3(i) + \gamma \phi_4(i) + \mu \phi_5(i) + \sigma \phi_6(i) + \frac{\delta}{\phi_7(i)} \quad (31)$$

in which the cost function is represented by J and in its formulation, H_p defines is the prediction horizon, n represents a discrete time variable and where $\lambda, \beta, \alpha, \gamma, \mu, \sigma$ and δ are weights for the objectives ϕ_i , for $i = 1, \dots, 7$. Each objective has a unique characteristic that will influence the behaviour of the wastewater system depending on the weights chosen for this objective.

The objective ϕ_1 , in Equation (32), is used to homogeneously store the volume within all tank structures in the sewage network.

$$\phi_1(n) = \sum_{i=1}^N \left(V_i(n) - \frac{V_{i_{max}}}{\sum_{j=1}^N V_{j_{max}}} \sum_{k=1}^N V_k(n) \right)^2 \quad (32)$$

where V is the volume in the tank, V_{max} the maximum volume in the tank, $i = 1, \dots, N$ is the number of tanks in the network and $j = 1, \dots, N_p$ the number of pipes.

ϕ_2 , in Equation (33), keeps the flow towards the wastewater treatment plant near the operating reference value defined by y_{ref} .

$$\phi_2(n) = \left(y_{ref}(n) - \sum_{i \in N_k^*} \text{Out}_i(n - d_{i,k}) \right)^2 \quad (33)$$

where Out is the outflow from each tank, $d_{i,k}$ represents the transport time of the i^{th} tank to the destination tank k and N_k^* are all tanks draining in the destination k .

ϕ_3 , in Equation (34), is used to minimise the overflow, where NL is assigned as a negative number to keep the objective linear.

$$\phi_3(n) = \sum_{i=1}^N (\text{Ov}_i(n) - NL)^2 \quad (34)$$

where Ov is the overflow volume.

ϕ_4 , in Equation (35), is used to minimise the load, e.g., reduce the mass in overflow events.

$$\phi_4(n) = \sum_{i=1}^N (C_i(n) \text{Ov}_i(n) - NL)^2 \quad (35)$$

where C is the concentration in the tank.

ϕ_5 , in Equation (36), is to minimise the uncertainty in the load. This objective is interesting in the case where multiple tanks are overflowing and in this scenario, the controller will prefer to let tanks with less uncertainty overflow.

$$\phi_5(n) = \sum_{i=1}^N (U_i(n) - NL)^2 \quad (36)$$

where U is the uncertainty related to the concentration within the tank.

ϕ_6 , in Equation (37), is similar to objective ϕ_1 , the controller uses this objective to balance the mass over the sewage system's structures homogeneously.

$$\phi_6(n) = \sum_{i=1}^N \left(C_i(n)V_i(n) - \frac{V_{i_{max}}}{\sum_{j=1}^N V_{j_{max}}} \sum_{k=1}^N C_k(n)V_k(n) \right)^2 \quad (37)$$

ϕ_7 , in Equation (38), is used to maximise the load arriving the wastewater treatment plant.

$$\phi_7(n) = \left(\sum_{i \in N_k^*} C_i(n - d_{i,k}) Out_i(n - d_{i,k}) \right)^2 \quad (38)$$

The final minimisation problem is given by Equation (31) subject to the constraints in Equation (39):

$$\begin{aligned} 0 &\leq V_i(n) \leq V_{i_{max}} \\ 0 &\leq Out_i(n) \leq Out_{i_{max}} \\ O_v(n) &\geq 0 \\ C_i(n) &\geq 0 \\ U_i(n) &\geq 0 \\ Q_{pipe_j}(n) &= \sum_{i \in N_j^*} Out_k(n - d_{i,k}) \leq Q_{pipe_{j_{max}}} \end{aligned} \quad (39)$$

where Q_{pipe} is the flow in a pipe and $Q_{pipe_{max}}$ is the maximum flow in a given pipe.

The method uses a simplified model as described in Section 2.2.1.

The pollution-based MPC reduced the mass overflowed compared to a volume-based approach [48]. This method is advantageous mainly due to the use of uncertainty, which is often disregarded during the development of a RTC system. The advantage is that, when considering uncertainty, it is assumed that there is an error related to the measurement or estimation of concentration, which is likely to occur. Therefore, the method is more robust to errors and performance is improved. However, this increased robustness comes at the cost of higher complexity and computational efforts, coupled with the need of a concentration model that requires a non-linear optimisation algorithm, increasing the complexity even further.

3.2.3. Multi-objective optimisation using genetic algorithms

As mentioned in Section 2, wastewater is composed of numerous pollutants and the quality of the water depends on the combination of these pollutants. The method presented by [62] is a pollution-based dynamic control algorithm for a sewer network using a multi-objective optimisation which considers different pollutants within the formulation of the optimisation problem. The author discusses the possible solutions for solving the multi-objective optimisation and selects a genetic algorithm-based optimisation due to the complexity of the problem.

Genetic algorithms (GA) are an heuristic method for solving optimisation problems [17] by mimicking the search for the optimum through biological evolution. In each time step, the solutions, also known as a population of solutions, are updated. Genetic algorithms have basic features to mimic evolutionary behaviour, such as chromosome representation, fitness selection and biological-inspired operators [41]. The overall optimisation problem is similar to Equation (25), however the objective is separated as in Equation (40):

$$\min J(n) = J_1(n) + J_2(n) \quad (40)$$

The cost J_1 is related to the minimisation of the pollutant load discharged from a sewer chamber, as can be seen in Equation (41):

$$J_1 = \sum_{i=1}^{n_{CSO}} EQI_i \quad (41)$$

where n_{CSO} is the number of CSOs and EQI is the effluent quality index given by Equation (12).

The second objective J_2 , shown in Equation (42), is used to minimise the cost of the treatment plant while also reducing operational costs from pumps.

$$J_2 = C_T + C_P \quad (42)$$

where C_T is the wastewater treatment plant cost and C_P is the operational cost of pumps in the sewer system. The calculation of both C_T and C_P are presented in [63].

The method uses a simplified model as described in Section 2.2.3.

The author concludes that the genetic algorithm with the given objectives is capable of minimising the pollution load while reducing pumping costs, which is an unique approach for real time controllers, most methods are more concerned with volume reduction without actually measuring any costs. The algorithm also produced varied temporally and spatially dynamic control

Table 4
Papers on optimisation-based RTC methods with pollution aspects.

Authors	Year	Type of Model	Type of Optimisation	Description	Type of Pollutant
[65]	1999	Model-free	MPC + GA	Nonlinear MPC using Genetic Algorithm for calculating optimal solution applied to the sewer system, treatment plant and receiving water with the aim to achieve minimum effects of pollution.	DO/COD
[28]	2008	Simplified model	GA	Multi-objective quality control of urban wastewater system using NSGA-II genetic algorithm. The algorithm evaluates the performance of the system directly with regard to receiving water quality indicators, such as for ammonia, phosphorus and dissolved oxygen.	Multiple pollutants
[66]	2010	Simplified model	MPC + Fuzzy	Investigation on fuzzy decision making for multi-criteria optimisation of sewer networks and sewage treatment works regarding the balance of ecologic and economic criteria.	Global indicator
[90]	2014	Simplified model	MPC	Global water-quality optimisation with dynamic cost functions that aims to reduce the global overflow risk across the catchment. The algorithm uses hydrological models with indirect water quality considerations to improve the system performance by dynamically changing the prioritisation of discharges at different CSO structures.	TSS
[89]	2014	Simplified model	Ant Colony	Optimisation-based method designed to improve pollution load management using an ant colony algorithm. The objective is to empty retention tanks using volume and water quality data and discharge the flow to the WWTP while also avoiding exceeding the maximum capacity of the plant.	Multiple pollutants
[78]	2018	Simplified model	MPC	Model predictive control for TSS minimisation in the framework of cyber-physical systems.	TSS
[85]	2020	Model-free	Market-based	Real-time load-balancing algorithm to control distributed storage assets in the collection system, to improve wet-weather flows and water quality at a receiving point.	TSS
[67]	2021	Simplified model	MPC	Integrated multi-software pollution-based control architecture using model predictive control. It proposes a methodology for a closed-loop application between the RTC module and the wastewater system for a water quality application.	TSS
[88]	2022	Simplified model	GA	Genetic algorithm optimisation with cost functions related to the volume of overflow and the overflow quality index for a sewer network, which indicates a mass of polluting units per day. The study compares five optimal control strategies using these metrics.	Multiple pollutants
[12]	2023	Simplified model	Decentralised MPC + Fuzzy	Distributed model predictive control algorithm including fuzzy negotiation among subsystems and a dynamic set-point generation method, applied to a simulated sewerage network using a hydraulic and water quality model for prediction.	TSS

settings, obtained depending on regulatory requirements and the financial and environmental situation of the sewer management agency [62]. The implementation of the proposed method is simple, however the calculation of the wastewater treatment plant and pumping cost is not trivial and might lead to uncertainties. However, it is mentioned in the paper that, although GA methods have a rapid convergence, using a genetic algorithm introduces a high computational cost that scales with the size of the sewage network and more efforts are required for real applications.

3.2.4. Section summary

In Table 4 several papers are shown with more implementations of pollution and optimisation-based RTC. The application of optimisation-based RTC is appealing for the efficient load management of urban drainage system. The high computational cost for solving optimisation problems can be a problem for network with non-linear dynamics and high number of decision variables, although this issue is diminished by advancements in computer performance.

3.3. Reinforcement learning-based RTC

Reinforcement Learning (RL) is a machine learning training method in which the main idea is to mimic the human behaviour of learning through environmental interactions. A RL problem involves a trial-and-error situation where the algorithm learns what to

do and derive an action of which the numerical reward from taking this action is maximised. By trying each action, the learner finds the action with most reward and influence later inputs of the system, which is why RL is a closed-loop algorithm [81].

The RL method is composed of an agent that takes actions, known as a , to interact with its environment expecting a reward, called r . During each learning step, the agent receives the state s of the environment and is rewarded depending by the quality of its actions to follow certain objectives, which can be a positive or negative reward [9]. Finally, the objective of the RL is to maximise the return function G_t as below in Equation (43):

$$G_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \dots = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \quad (43)$$

where $r_t = r(s_t, a_t, s_{t+1})$ and $\gamma \in [0, 1]$ is a discount factor weighting the importance of short-term and long-term reward that depends on the current action a_t , the current and next states s_t and s_{t+1} , respectively [81].

Although research in this field is growing significantly for storm-water systems regarding flood mitigation and peak flow reduction, such as the works in [9] and [53], a reinforcement learning RTC with the goal of improving water quality is still in its infancy.

3.3.1. Reinforcement learning-based RTC to improve water quality

The research is a novel method using a pollution-based approach proposed by [10] using a reinforcement learning algorithm for developing the RTC. The objective of the controller is to mitigate flooding and improve water quality using a RL-based RTC, where the environment used for the learning process is executed using the software SWMM [69]. The model in SWMM provides the current states to the RL, which are:

$$s_t = \begin{cases} \text{Current depth} \\ \text{Outflow} \\ \text{Concentration of TSS} \\ \text{Current valve positions} \\ \text{Sum of 24 h rain forecast} \\ \text{Mean value of the 24 h tide forecast} \end{cases}$$

The action space (A) of the agent is manipulating the valve to any desired value. The reward (r) is calculated by how well the objectives defined are met by the agent after taking any action within the action space, where the objectives are flood mitigation and pollutant reduction.

The author based the reinforcement learning algorithm on Deep Deterministic Policy Gradients (DDPG) [44] using a deep feed-forward neural network. The DDPG responsible to calculate the control actions (valve opening) over a continuous space while learning to control the system following the desired objectives maximising the rewards. It is presented a RL-based RTC for a sewage system with 2 sub catchments, called Pond 1 and Pond 2. Three agents were created and trained using the DDPG algorithm with the following reward system:

- Agent 1, r_{ag1} , is rewarded for minimising total flooding and keeping target pond depths, therefore it is not used for water quality. The reward function is as follows in Equation (44):

$$r_{ag1} = \begin{cases} -\sum \text{Flooding}[\text{system}, \text{Pond1} * 1000, \text{Pond2}] & F \geq \delta \\ -(|\text{Pond1}_{depth} - \tau_1| + |\text{Pond2}_{depth} - \tau_2|) & F < \delta \end{cases} \quad (44)$$

where $\text{Flooding}[\text{system}]$ is the total flooding volume, $\text{Flooding}[\text{Pond1}]$ is the flooding rate at Pond 1 and $\text{Flooding}[\text{Pond2}]$ is the reward for having Pond 2 depth above a level that causes upstream flooding. F and δ are the total rainfall in a 24 h window and the rainfall threshold, respectively. Finally, τ_1 and τ_2 is the reference value for the depth of Pond 1 and Pond 2, respectively.

- In the Agent 2, r_{ag2} , is added water quality rewards, which results in it being rewarded for minimising total flooding, keeping target pond depths and minimising the outflow of TSS from the ponds as in Equation (45):

$$r_{ag2} = \begin{cases} -\sum \text{Flooding}[\text{system}, \text{Pond1} * 1000, \text{Pond2}] + TSS[\text{Valve1}, \text{Valve2}] & F \geq \delta \\ -(|\text{Pond1}_{depth} - \tau_1| + |\text{Pond2}_{depth} - \tau_2|) + TSS[\text{Valve1}, \text{Valve2}] + \text{Flooding}[\text{system}/35000] & F < \delta \end{cases} \quad (45)$$

where $TSS[\text{Valve1}, \text{Valve2}]$ is the total TSS load of the controlled valves.

- Agent 3: is used as balance between Agent 1 and Agent 2. The agent is set by using the weights and memory from the trained neural network of Agent 1 as an initialisation for the reward function r_{agent2} from Agent 2, functioning as a pre-training of Agent 3.

Bowes discusses the performance of the controller comparing the results using a RL-based RTC with each of the agents. The increased complexity of Agent 2 compared to Agent 1, due to the extra control objective for minimising the TSS load, revealed that it is harder for a single agent to effectively learn all control policies. Agent 2 was able to effectively learn a policy for minimising the

TSS load, however Agent 1 had a better performance for flood mitigation, which indicates that Agent 2 compromised by increasing the flooding. The pre-training performed by Agent 3 using Agent 1's information into Agent 2's reward function resulted in a balanced performance between agents 1 and 2. [10] mention that correctly weighting the RL is highly important to avoid reward gaming, which happens when agent's actions causes undesired behaviours in the environment.

The method uses a model based on neural networks and, therefore, inherently uses a data-driven modelling technique. The author suggests that actual physical or numerical stormwater models could be incorporated into the neural network, but further evaluation is recommended.

3.3.2. Section summary

As mentioned, the use of RL for pollution-based RTC is a recent development. Mullapudi [53] discusses that there are challenges associated with RL, such as formulating rewards, choosing function approximators and deciding the complexity of the control problem. It is stated that for many fields the control is highly sensitive to the formulation of the reward function and requires both expertise and an element of subjectivity for a proper implementation. Moreover, to learn all possible relationships between rewards and control decisions becomes a complex task with high computational demands. Although the paper's statements are related to stormwater system control, the same conclusions should apply to wastewater network control but may also explain the lack of studies in the field.

4. Conclusion

The research in pollution-based real time controllers is evolving and the future for its application in sewer systems seems promising. Pollution-based Real Time Control theoretically has the potential for out-performing volume-based systems due to the possibility to explicitly consider pollution in the formulation of the controller. This creates the opportunity to directly affect the quality through the sewage network. P-RTC is also a cost-effective option for improving receiving water quality, reducing the load of the discharged wastewater and guarantee a better functioning of the wastewater treatment plant. However, many obstacles still remain:

- **Nonlinear dynamics are added to the problem when considering concentration within the control formulation**, which increase the challenge of defining models, increase the complexity of the control system and requiring non-linear optimisation. This results in a higher computational cost to calculate the controlled variable in each sampling interval. Several solutions were introduced, such as using a simplified non-linear model based with mass balance equations or data-driven models that, with enough information about the dynamics in the data, will be able to represent the system consistently;
- **Uncertainties are present from measuring devices and estimation techniques** used for real time monitoring of concentration in the wastewater distribution, which increase errors for calculating control actions. The research on uncertainty for control system methods based on pollution is still limited and it is in need of further investigation to understand the impacts in the performance of the controller. Until there is clarity on this aspect, the performance of a P-RTC cannot be properly assessed.

This review started by detailing several modelling methods with pollution intrinsically in its formulation and for control purposes. There is a lack of consensus in the literature of which modelling technique to use, in general. For prediction, although many models are proposed, researchers are still searching for a benchmark effective model that represents the sewage systems with complex dynamics. Many authors use simple formulations for reducing complexity to use in optimisation techniques, but the main issue is the lack of predictability of some dynamics, such as infiltration and first flush effect. Machine learning techniques are a good alternative for creating models with complex dynamics without actually knowing the behaviour of the network, however these algorithms rely on data, which are directly influenced by uncertainty and might result in erroneous and unwanted behaviours by the model.

The P-RTC methods in general are great for improving water quality management. Integrating pollutant prediction models with P-RTC methods has a potential to improve the efficiency on sewage management and comply with governmental water quality targets.

Regarding the control techniques, the research showed a predominance of rule-based and optimisation-based algorithms, however reinforced learning is a promising subject that just recently started getting visibility through the work of [10]. The most simple approach is the RB-RTC, which is defined by a set of rules that are employed to manipulate the sewage system towards a desired quality behaviour. The simplicity of the method is the main appeal for its use, requiring little to zero intervention in the wastewater network system, while also being intuitive for operators. Although many of the RB-RTC implementations are integrated with a volume-based optimisation, usually they are local-based, which means the controllers do not observe the complete network, only local structures individually. Besides that, understanding the dynamics of the sewage system is essential for setting the rules, any mistake will deviate from the desired behaviour and increasing the number of rules will reduce the intuitiveness of the algorithm.

The optimisation-based RTC literature showed that this category is mainly comprised by methods based on model predictive control and genetic algorithms. The similarities are in the implementation of objective functions that determine the control actions and the use of an optimisation method, however each method solves the problem differently. Optimisation-based methods are, more often than not, global, as opposite to RB-RTC, and the control actions are optimal, i.e., the controller calculates the best optimal action. Nonetheless, these methods are complex and demand, simultaneously, knowledge from operators to properly configure the controller and higher computational cost from hardware. A third alternative reviewed is reinforcement learning-based methods, which are algorithms based on mimicking natural evolution and predicting the behaviour through a learning process. The RL-based

RTC is new and more developments are needed to understand the performance compared to other methods, however it showed as a promising non-optimisation dependant alternative for implementing global real time control.

The literature in pollution-based RTC methods focused its efforts in presenting novel controllers and comparing the performance with previous implementations, however there is still much to be researched. For future research topics, the authors suggest:

- An in-depth water quality control process identification study is missing, complementary to the research in [91], which studies and proposes the identification of variables such as the volume, flow rate, velocity and depth, in a pilot sewage network using the conservation of mass and momentum;
- The inflow concentration C_{in} is one of the most important variables for an optimal control of the wastewater network and it is uncertain. Research studying the uncertainty in C_{in} and how to handle it in RTC systems is limited [55]. Robustness is also important for keeping stability, but research in this field is not existing for P-RTC methods;
- There are some studies in stochastic volume-based RTC ([82] and [33]) and drinking water networks ([24] and [18]), however using this framework within pollution-based RTC is presently missing and this could be an interesting research field;
- Data-driven models for prediction in SUDS are a well-known topic in literature and the application of these models in volume-based RTC can be observed in [5], [23] and [45]. However, the research of data-driven methods for water quality prediction used in P-RTC is still in its infancy, thus being an appealing research topic for future research;
- Further developments on reinforcement learning RTC that is based solely on the research done by [10], such as different reward shapes, applying a multi-objective RL P-RTC to deal with both quantity and quality targets or study RL optimisation for different pollutant types;
- More studies in real applications of pollution-based RTC.

The potential of pollution-based RTC to have an impact on water quality may be considered to be illustrated by the increase in research efforts. Also realistically, urban drainage systems are ready for implementation of these innovative methods because they are increasingly outfitted with modern PLC and SCADA systems. Hence, future researchers need to focus on real implementation of pollution-based methods and demonstrate the benefits.

CRediT authorship contribution statement

Rodrigo da Silva Gesser: Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Holger Voos:** Supervision, Resources, Funding acquisition, Formal analysis. **Alex Cornelissen:** Writing – review & editing, Validation, Supervision, Resources, Project administration. **Georges Schutz:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Rodrigo da Silva Gesser reports financial support was provided by Fonds National de la Recherche - FNR, grant reference 17139914. Rodrigo da Silva Gesser reports financial support and writing assistance were provided by University of Luxembourg, grant reference 17139914. Georges Schutz reports a relationship with RTC4Water that includes: funding grants and non-financial support.

Data availability

No data was used for the research described in the article.

Acknowledgements

This research was funded in whole, or in part, by the Luxembourg National Research Fund (FNR), grant reference 17139914. For the purpose of open access, and in fulfilment of the obligations arising from the grant agreement, the author has applied a Creative Commons Attribution 4.0 International (CC BY 4.0) license to any Author Accepted Manuscript version arising from this submission. The authors would like to thank RTC4Water, the Interdisciplinary Centre for Security, Reliability and Trust (SnT) and the Faculty of Science, Technology and Medicine (FSTM) and Dept. of Engineering from University of Luxembourg for supporting and funding the research.

References

- [1] R.R.A. Al-Ani, B.H.K. Al-Obaidi, Prediction of sediment accumulation model for trunk sewer using multiple linear regression and neural network techniques, *Civ. Eng. J.* 5 (2019) 82, <https://doi.org/10.28991/cej-2019-03091227>.
- [2] S. Arthur, R.M. Ashley, The influence of near bed solids transport on first foul flush in combined sewers, *Water Sci. Technol.* 37 (1998), [https://doi.org/10.1016/s0273-1223\(97\)00762-2](https://doi.org/10.1016/s0273-1223(97)00762-2).
- [3] P.M. Bach, W. Rauch, P.S. Mikkelsen, D.T. McCarthy, A. Deletic, A critical review of integrated urban water modelling – urban drainage and beyond, *Environ. Model. Softw.* 54 (2014) 88–107, <https://doi.org/10.1016/j.envsoft.2013.12.018>.

- [4] A. Bachmann-Machnik, Y. Brüning, A.E. Bakhshipour, M. Krauss, U. Dittmer, Evaluation of combined sewer system operation strategies based on highly resolved online data, *Water* 13 (2021) 751, <https://doi.org/10.3390/w13060751>.
- [5] K.M. Balla, J.D. Bendtsen, C. Schou, C.S. Kallesoe, C. Ocampo-Martinez, A learning-based approach towards the data-driven predictive control of combined wastewater networks – an experimental study, *Water Res.* 221 (2022) 118782, <https://doi.org/10.1016/j.watres.2022.118782>, <https://www.sciencedirect.com/science/article/pii/S0043135422007357>.
- [6] K.M. Balla, C. Schou, J.D. Bendtsen, C. Ocampo-Martinez, C.S. Kallesoe, A nonlinear predictive control approach for urban drainage networks using data-driven models and moving horizon estimation, *IEEE Trans. Control Syst. Technol.* 30 (2022) 2147–2162, <https://doi.org/10.1109/tcst.2021.3137712>.
- [7] L.R. Ben, C. Sun, R.G. Palma, B.J. Duran, J. Meseguer, G. Cembrano, V. Puig, A feedback simulation procedure for real-time control of urban drainage systems, *IFAC-PapersOnLine* 52 (2019) 101–106, <https://doi.org/10.1016/j.ifacol.2019.11.016>.
- [8] W. Bourgeois, J.E. Burgess, R.M. Stuetz, On-line monitoring of wastewater quality: a review, *J. Chem. Technol. Biotechnol.* 76 (2001) 337–348, <https://doi.org/10.1002/jctb.393>.
- [9] B.D. Bowes, A. Tavakoli, C. Wang, A. Heydarian, M. Behl, P.A. Beling, J.L. Goodall, Flood mitigation in coastal urban catchments using real-time stormwater infrastructure control and reinforcement learning, *J. Hydroinform.* 23 (2020) 529–547, <https://doi.org/10.2166/hydro.2020.080>.
- [10] B.D. Bowes, C. Wang, M.B. Ercan, T.B. Culver, P.A. Beling, J.L. Goodall, Reinforcement learning-based real-time control of coastal urban stormwater systems to mitigate flooding and improve water quality, *Environmental Science: Water Research & Technology* 8 (2022) 2065–2086, <https://doi.org/10.1039/d1ew00582k>.
- [11] P.J. Braspenning, F. Thuijsman, A.J.M.M. Weijters (Eds.), *Artificial Neural Networks*, Springer, Berlin Heidelberg, 1995.
- [12] A. Cembellín, M. Francisco, P. Vega, Optimal operation of a benchmark simulation model for sewer networks using a qualitative distributed model predictive control algorithm, *Processes* 11 (2023), <https://doi.org/10.3390/pr11051528>, <https://www.mdpi.com/2227-9717/11/5/1528>.
- [13] G. Cembrano, Optimal control of urban drainage systems. a case study, *Control Eng. Pract.* 12 (2004) 1–9, [https://doi.org/10.1016/s0967-0661\(02\)00280-0](https://doi.org/10.1016/s0967-0661(02)00280-0).
- [14] M.R. di Cicco, A. Masiello, A. Spagnuolo, C. Vetromile, L. Borea, G. Giannella, M. Iovinella, C. Lubritto, Real-time monitoring and static data analysis to assess energetic and environmental performances in the wastewater sector: a case study, *Energies* 14 (2021) 6948, <https://doi.org/10.3390/en14216948>.
- [15] I. Corporation, *IBM SPSS Neural Networks* 21, 2013.
- [16] P. van Daal, G. Gruber, J. Langeveld, D. Muschalla, F. Clemens, Performance evaluation of real time control in urban wastewater systems in practice: review and perspective, *Environ. Model. Softw.* 95 (2017) 90–101, <https://doi.org/10.1016/j.envsoft.2017.06.015>.
- [17] L. Davis (Ed.), *Handbook of Genetic Algorithms*, Van Nostrand Reinhold, 1991.
- [18] A. Dhar, B. Datta, Chance constrained water quality management model for reservoir systems, *ISH J. Hydraul. Eng.* 12 (2006) 39–48, <https://doi.org/10.1080/09715010.2006.10514848>.
- [19] W. Duan, F.J. del Campo, M. Gich, C. Fernández-Sánchez, In-field one-step measurement of dissolved chemical oxygen demand with an integrated screen-printed electrochemical sensor, *Sens. Actuators B, Chem.* 369 (2022) 132304, <https://doi.org/10.1016/j.snb.2022.132304>.
- [20] S. Duchesne, A. Mailhot, J.P. Villeneuve, Global predictive real-time control of sewers allowing surcharged flows, *J. Environ. Eng.* 130 (2004) 526–534, [https://doi.org/10.1061/\(asce\)0733-9372\(2004\)130:5\(526\)](https://doi.org/10.1061/(asce)0733-9372(2004)130:5(526)).
- [21] I. Ebtehaj, H. Bonakdari, A.H. Zaji, A new hybrid decision tree method based on two artificial neural networks for predicting sediment transport in clean pipes, *Alex. Eng. J.* 57 (2018) 1783–1795, <https://doi.org/10.1016/j.aej.2017.05.021>.
- [22] S. Eggimann, L. Mutzner, O. Wani, M.Y. Schneider, D. Spuhler, M.M. de Vitry, P. Beutler, M. Maurer, The potential of knowing more: a review of data-driven urban water management, *Environ. Sci. Technol.* 51 (2017) 2538–2553, <https://doi.org/10.1021/acs.est.6b04267>.
- [23] K. El Ghazouli, J. El Khatabi, A. Souhli, I. Shahrou, Model predictive control based on artificial intelligence and epa-swmm model to reduce csos impacts in sewer systems, *Water Sci. Technol.* 85 (2021) 398–408, <https://doi.org/10.2166/wst.2021.511>, <https://iwaponline.com/wst/article-pdf/85/1/398/986065/wst085010398.pdf>.
- [24] F. Fiedler, A. Cominola, S. Lucia, Economic nonlinear predictive control of water distribution networks based on surrogate modeling and automatic clustering, *IFAC-PapersOnLine* 53 (2020) 16636–16643, <https://doi.org/10.1016/j.ifacol.2020.12.793>.
- [25] D. Fiorelli, G. Schutz, Real-time control of a sewer network using a multi-goal objective function, in: *2009 17th Mediterranean Conference on Control and Automation, IEEE*, 2009.
- [26] D. Fiorelli, G. Schutz, K. Klepiszewski, M. Regneri, S. Seiffert, Optimised real time operation of a sewer network using a multi-goal objective function, *Urban Water J.* 10 (2013) 342–353, <https://doi.org/10.1080/1573062x.2013.806560>.
- [27] J. Foley, Z. Yuan, P. Lant, Dissolved methane in rising main sewer systems: field measurements and simple model development for estimating greenhouse gas emissions, *Water Sci. Technol.* 60 (2009) 2963–2971, <https://doi.org/10.2166/wst.2009.718>.
- [28] G. Fu, D. Butler, S.T. Khu, Multiple objective optimal control of integrated urban wastewater systems, *Environ. Model. Softw.* 23 (2008) 225–234, <https://doi.org/10.1016/j.envsoft.2007.06.003>.
- [29] L. Fuchs, T. Beenenken, Development and implementation of a real-time control strategy for the sewer system of the city of Vienna, *Water Sci. Technol.* 52 (2005) 187–194, <https://doi.org/10.2166/wst.2005.0133>.
- [30] E. Gaborit, D. Muschalla, B. Vallet, P.A. Vanrolleghem, F. Antil, Improving the performance of stormwater detention basins by real-time control using rainfall forecasts, *Urban Water J.* 10 (2013) 230–246, <https://doi.org/10.1080/1573062x.2012.726229>.
- [31] Y. Gao, X. Shi, X. Jin, X.C. Wang, P. Jin, A critical review of wastewater quality variation and in-sewer processes during conveyance in sewer systems, *Water Res.* 228 (2023) 119398, <https://doi.org/10.1016/j.watres.2022.119398>.
- [32] L. García, J. Barreiro-Gomez, E. Escobar, D. Téllez, N. Quijano, C. Ocampo-Martinez, Modeling and real-time control of urban drainage systems: a review, *Adv. Water Resour.* 85 (2015) 120–132, <https://doi.org/10.1016/j.advwatres.2015.08.007>.
- [33] J. Grosso, C. Ocampo-Martínez, V. Puig, B. Joseph, Chance-constrained model predictive control for drinking water networks, *J. Process Control* 24 (2014) 504–516, <https://doi.org/10.1016/j.jprocont.2014.01.010>.
- [34] D.W. Hook, S.J. Porter, C. Herzog, Dimensions: building context for search and evaluation, *Front. Res. Metr. Anal.* 3 (2018) 23, <https://doi.org/10.3389/frma.2018.00023>, <https://app.dimensions.ai/details/publication/pub.1106289502>, <https://www.frontiersin.org/articles/10.3389/frma.2018.00023/pdf>.
- [35] H. Hoppe, S. Messmann, A. Giga, H. Gruening, A real-time control strategy for separation of highly polluted storm water based on UV–vis online measurements – from theory to operation, *Water Sci. Technol.* 63 (2011) 2287–2293, <https://doi.org/10.2166/wst.2011.164>.
- [36] O. Intrator, N. Intrator, Interpreting neural-network results: a simulation study, *Comput. Stat. Data Anal.* 37 (2001) 373–393, [https://doi.org/10.1016/s0167-9473\(01\)00016-0](https://doi.org/10.1016/s0167-9473(01)00016-0).
- [37] K. Irvine, M.C. Rossi, S. Vermette, J. Bakert, K. Kleinfelder, Illicit discharge detection and elimination: low cost options for source identification and trackdown in stormwater systems, *Urban Water J.* 8 (2011) 379–395, <https://doi.org/10.1080/1573062x.2011.630095>.
- [38] Y. Jia, F. Zheng, H.R. Maier, A. Ostfeld, E. Creaco, D. Savic, J. Langeveld, Z. Kapelan, Water quality modeling in sewer networks: review and future research directions, *Water Res.* 202 (2021) 117419, <https://doi.org/10.1016/j.watres.2021.117419>.
- [39] G. Jiang, J. Keller, P.L. Bond, Z. Yuan, Predicting concrete corrosion of sewers using artificial neural network, *Water Res.* 92 (2016) 52–60, <https://doi.org/10.1016/j.watres.2016.01.029>.
- [40] Y. Jiang, C. Li, L. Sun, D. Guo, Y. Zhang, W. Wang, A deep learning algorithm for multi-source data fusion to predict water quality of urban sewer networks, *J. Clean. Prod.* 318 (2021) 128533, <https://doi.org/10.1016/j.jclepro.2021.128533>.
- [41] S. Katoch, S.S. Chauhan, V. Kumar, A review on genetic algorithm: past, present, and future, *Multimed. Tools Appl.* 80 (2020) 8091–8126, <https://doi.org/10.1007/s11042-020-10139-6>.

- [42] M. Kim, A.S. Rao, C. Yoo, Dual optimization strategy for n and p removal in a biological wastewater treatment plant, *Ind. Eng. Chem. Res.* 48 (2009) 6363–6371, <https://doi.org/10.1021/ie801689t>.
- [43] S. Konishi, *Introduction to Multivariate Analysis*, Chapman and Hall/CRC, 2014.
- [44] T.P. Lillicrap, J.J. Hunt, A. Pritzel, N. Heess, T. Erez, Y. Tassa, D. Silver, D. Wierstra, Continuous control with deep reinforcement learning, <https://doi.org/10.48550/ARXIV.1509.02971>, 2015.
- [45] T. Liu, J.E. Ramirez-Marquez, S.C. Jagupilla, V. Prigiobbe, Combining a statistical model with machine learning to predict groundwater flooding (or infiltration) into sewer networks, *J. Hydrol.* 603 (2021) 126916, <https://doi.org/10.1016/j.jhydrol.2021.126916>, <https://www.sciencedirect.com/science/article/pii/S0022169421009665>.
- [46] S. Ma, T. Zayed, J. Xing, Y. Shao, A state-of-the-art review for the prediction of overflow in urban sewer systems, *J. Clean. Prod.* 434 (2024) 139923, <https://doi.org/10.1016/j.jclepro.2023.139923>.
- [47] J. Maciejowski, *Predictive Control with Constraints*, Prentice Hall, 2002.
- [48] M. Mahmoodian, O. Delmont, G. Schutz, Pollution-based model predictive control of combined sewer networks, considering uncertainty propagation, *International Journal of Sustainable Development and Planning* 12 (2017) 98–111, <https://doi.org/10.2495/sdp-v12-n1-98-111>.
- [49] C. Man, C.W. Tsai, Stochastic partial differential equation-based model for suspended sediment transport in surface water flows, *J. Eng. Mech.* 133 (2007) 422–430, [https://doi.org/10.1061/\(asce\)0733-9399\(2007\)133:4\(422\)](https://doi.org/10.1061/(asce)0733-9399(2007)133:4(422)).
- [50] P. Martin, C. Barranquero, J. Sanchez, G. Gestoso, J. Andres, D. Refoyo, C. Garcia, A. Fernandez, R. Lopez, Opennn: open neural network library, <https://www.opennn.net/>, 2021.
- [51] A.L. Mollerup, P.S. Mikkelsen, D. Thornberg, G. Sin, Controlling sewer systems – a critical review based on systems in three EU cities, *Urban Water J.* 14 (2016) 435–442, <https://doi.org/10.1080/1573062x.2016.1148183>.
- [52] V.M. Morales, J.C. Quijano, A. Schmidt, M.H. Garcia, Innovative framework to simulate the fate and transport of nonconservative constituents in urban combined sewer catchments, *Water Resour. Res.* 52 (2016) 9164–9181, <https://doi.org/10.1002/2016wr018807>.
- [53] A. Mullaipudi, M.J. Lewis, C.L. Gruden, B. Kerkez, Deep reinforcement learning for the real time control of stormwater systems, *Adv. Water Resour.* 140 (2020) 103600, <https://doi.org/10.1016/j.advwatres.2020.103600>.
- [54] C. Ocampo-Martinez, *Model Predictive Control of Wastewater Systems*, Springer, London, 2010.
- [55] J. Oh, M. Bartos, Model predictive control of stormwater basins coupled with real-time data assimilation enhances flood and pollution control under uncertainty, *Water Res.* 235 (2023) 119825, <https://doi.org/10.1016/j.watres.2023.119825>, <https://www.sciencedirect.com/science/article/pii/S0043135423002609>.
- [56] J. Oh, C.W. Tsai, A stochastic multivariate framework for modeling movement of discrete sediment particles in open channel flows, *Stoch. Environ. Res. Risk Assess.* 32 (2017) 385–399, <https://doi.org/10.1007/s00477-017-1410-3>.
- [57] J. Pan, T. Leng, Y. Liu, Shifosi reservoir water environmental assessment based on grey clustering, *Adv. Mater. Res.* 610–613 (2012) 857–861, <https://doi.org/10.4028/www.scientific.net/amr.610-613.857>.
- [58] O. Panasiuk, A. Hedström, J. Marsalek, R.M. Ashley, M. Viklander, Contamination of stormwater by wastewater: a review of detection methods, *J. Environ. Manag.* 152 (2015) 241–250, <https://doi.org/10.1016/j.jenvman.2015.01.050>.
- [59] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, *J. Mach. Learn. Res.* 12 (2011) 2825–2830.
- [60] F. Piadeh, K. Behzadian, A.M. Alani, A critical review of real-time modelling of flood forecasting in urban drainage systems, *J. Hydrol.* 607 (2022) 127476, <https://doi.org/10.1016/j.jhydrol.2022.127476>.
- [61] M. Pleau, H. Colas, P. Lavallee, G. Pelletier, R. Bonin, Global optimal real-time control of the Quebec urban drainage system, *Environ. Model. Softw.* 20 (2005) 401–413, <https://doi.org/10.1016/j.envsoft.2004.02.009>.
- [62] U. Rathnayake, A.F. Anwar, Dynamic control of urban sewer systems to reduce combined sewer overflows and their adverse impacts, *J. Hydrol.* 579 (2019) 124150, <https://doi.org/10.1016/j.jhydrol.2019.124150>.
- [63] U.S. Rathnayake, T.T. Tanyimboh, Evolutionary multi-objective optimal control of combined sewer overflows, *Water Resour. Manag.* 29 (2015) 2715–2731, <https://doi.org/10.1007/s11269-015-0965-3>.
- [64] W. Rauch, J.L. Bertrand-Krajewski, P. Krebs, O. Mark, W. Schilling, M. Schütze, P. Vanrolleghem, Deterministic modelling of integrated urban drainage systems, *Water Sci. Technol.* 45 (2002) 81–94, <https://doi.org/10.2166/wst.2002.0059>.
- [65] W. Rauch, P. Harremoës, Genetic algorithms in real time control applied to minimize transient pollution from urban wastewater systems, *Water Res.* 33 (1999) 1265–1277, [https://doi.org/10.1016/s0043-1354\(98\)00304-2](https://doi.org/10.1016/s0043-1354(98)00304-2).
- [66] M. Regneri, K. Klepiszewski, M. Ostrowski, P. Vanrolleghem, Fuzzy decision making for multi-criteria optimization in integrated wastewater system management, 2010.
- [67] L. Romero, B. Joseph-Duran, C. Sun, J. Meseguer, G. Cembrano, R. Guasch, M. Martínez, E. Muñoz, V. Puig, An integrated software architecture for the pollution-based real-time control of urban drainage systems, *J. Hydroinform.* 23 (2021) 671–687, <https://doi.org/10.2166/hydro.2021.149>.
- [68] L. Rossi, V. Krejci, W. Rauch, S. Kreikenbaum, R. Fankhauser, W. Gujer, Stochastic modeling of total suspended solids (TSS) in urban areas during rain events, *Water Res.* 39 (2005) 4188–4196, <https://doi.org/10.1016/j.watres.2005.07.041>.
- [69] L.A. Rossman, et al., *Storm water management model user's manual, version 5.0*, National Risk Management Research Laboratory, Office of Research and ..., 2010.
- [70] P.M. Saco, P. Kumar, Kinematic dispersion in stream networks 1. coupling hydraulic and network geometry, *Water Resour. Res.* 38 (2002) 26–1–26–14, <https://doi.org/10.1029/2001wr000695>.
- [71] M.R. Schütze, D. Butler, M.B. Beck, *Modelling, Simulation and Control of Urban Wastewater Systems*, Springer, London, 2002.
- [72] S. Sharior, W. McDonald, A.J. Parolari, Improved reliability of stormwater detention basin performance through water quality data-informed real-time control, *J. Hydrol.* 573 (2019) 422–431, <https://doi.org/10.1016/j.jhydrol.2019.03.012>.
- [73] S. Shishegar, S. Duchesne, G. Pelletier, Optimization methods applied to stormwater management problems: a review, *Urban Water J.* 15 (2018) 276–286, <https://doi.org/10.1080/1573062x.2018.1439976>.
- [74] S. Shishegar, S. Duchesne, G. Pelletier, An integrated optimization and rule-based approach for predictive real time control of urban stormwater management systems, *J. Hydrol.* 577 (2019) 124000, <https://doi.org/10.1016/j.jhydrol.2019.124000>.
- [75] S. Shishegar, S. Duchesne, G. Pelletier, R. Ghorbani, A smart predictive framework for system-level stormwater management optimization, *J. Environ. Manag.* 278 (2021) 111505, <https://doi.org/10.1016/j.jenvman.2020.111505>.
- [76] A. Sikorska, D.D. Giudice, K. Banasik, J. Rieckermann, The value of streamflow data in improving TSS predictions – Bayesian multi-objective calibration, *J. Hydrol.* 530 (2015) 241–254, <https://doi.org/10.1016/j.jhydrol.2015.09.051>.
- [77] A.E. Sikorska, A. Scheidegger, K. Banasik, J. Rieckermann, Bayesian uncertainty assessment of flood predictions in ungauged urban basins for conceptual rainfall-runoff models, *Hydrol. Earth Syst. Sci.* 16 (2012) 1221–1236, <https://doi.org/10.5194/hess-16-1221-2012>.
- [78] C. Sun, G. Cembrano, V. Puig, J. Meseguer, Cyber-physical systems for real-time management in the urban water cycle, in: *2018 International Workshop on Cyber-Physical Systems for Smart Water Networks*, IEEE, 2018.
- [79] C. Sun, B. Joseph-Duran, T. Maruejols, G. Cembrano, J. Meseguer, V. Puig, X. Litrico, Real-time control-oriented quality modelling in combined urban drainage networks, *IFAC-PapersOnLine* 50 (2017) 3941–3946, <https://doi.org/10.1016/j.ifacol.2017.08.142>.
- [80] C. Sun, L. Romero, B. Joseph-Duran, J. Meseguer, E. Muñoz, R. Guasch, M. Martínez, V. Puig, G. Cembrano, Integrated pollution-based real-time control of sanitation systems, *J. Environ. Manag.* 269 (2020) 110798, <https://doi.org/10.1016/j.jenvman.2020.110798>.

- [81] R.S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction*, The MIT Press, 2014.
- [82] J.L. Svensen, C. Sun, G. Cembrano, V. Puig, Chance-constrained stochastic MPC of astlingen urban drainage benchmark network, *Control Eng. Pract.* 115 (2021) 104900, <https://doi.org/10.1016/j.conengprac.2021.104900>.
- [83] M.C. Ten Veldhuis, S. Tait, Data-driven urban drainage analysis: an alternative to hydrodynamic models, in: *Proceedings of the 12th International Conference on Urban Drainage*, Porto Alegre, Brazil 101, 2011, pp. 33–45.
- [84] J. Torres-Matallana, U. Leopold, K. Klepizewski, G. Heuvelink, EmiStatR: a simplified and scalable urban water quality model for simulation of combined sewer overflows, *Water* 10 (2018) 782, <https://doi.org/10.3390/w10060782>.
- [85] S.C. Troutman, N.G. Love, B. Kerkez, Balancing water quality and flows in combined sewer systems using real-time control, *Environmental Science: Water Research & Technology* 6 (2020) 1357–1369, <https://doi.org/10.1039/c9ew00882a>.
- [86] S.C. Troutman, N. Schambach, N.G. Love, B. Kerkez, An automated toolchain for the data-driven and dynamical modeling of combined sewer systems, *Water Res.* 126 (2017) 88–100, <https://doi.org/10.1016/j.watres.2017.08.065>.
- [87] P. Vanrolleghem, L. Benedetti, J. Meirlaen, Modelling and real-time control of the integrated urban wastewater system, *Environ. Model. Softw.* 20 (2005) 427–442, <https://doi.org/10.1016/j.envsoft.2004.02.004>.
- [88] I. Vasiliev, L. Luca, M. Barbu, R. Vilanova, S. Caraman, Optimal control strategy of a sewer network, *Water* 14 (2022) 1062, <https://doi.org/10.3390/w14071062>.
- [89] M. Verdager, N. Clara, O. Gutiérrez, M. Poch, Application of ant-colony-optimization algorithm for improved management of first flush effects in urban wastewater systems, *Sci. Total Environ.* 485–486 (2014) 143–152, <https://doi.org/10.1016/j.scitotenv.2014.02.140>.
- [90] L. Vezzaro, M. Christensen, C. Thirsing, M. Grum, P. Mikkelsen, Water quality-based real time control of integrated urban drainage systems: a preliminary study from Copenhagen, Denmark, *Proc. Eng.* 70 (2014) 1707–1716, <https://doi.org/10.1016/j.proeng.2014.02.188>.
- [91] Y. Wang, F. Haugen, Subspace system identification of a pilot tunnel system of a combined sewage system, *Modeling, Identification and Control: A Norwegian Research Bulletin* 44 (2023) 69–82, <https://doi.org/10.4173/mic.2023.2.3>.
- [92] Z. Wei, H. Shangguan, J. Zhan, R. Lin, X. Huang, L. Lu, H. Li, B. Du, G. Fan, Water quality-based double-gates control strategy for combined sewer overflows pollution control, *Water* 13 (2021) 529, <https://doi.org/10.3390/w13040529>.
- [93] G. Weinreich, W. Schilling, A. Birkely, T. Moland, Pollution based real time control strategies for combined sewer systems, *Water Sci. Technol.* 36 (1997) 331–336, <https://doi.org/10.2166/wst.1997.0688>.
- [94] T. Wells, R. Melchers, Modelling concrete deterioration in sewers using theory and field observations, *Cem. Concr. Res.* 77 (2015) 82–96, <https://doi.org/10.1016/j.cemconres.2015.07.003>.
- [95] J.A. van der Werf, Z. Kapelan, J. Langeveld, Towards the long term implementation of real time control of combined sewer systems: a review of performance and influencing factors, *Water Sci. Technol.* 85 (2022) 1295–1320, <https://doi.org/10.2166/wst.2022.038>.
- [96] W.D. Xu, M.J. Burns, F. Cherqui, T.D. Fletcher, Enhancing stormwater control measures using real-time control technology: a review, *Urban Water J.* 18 (2020) 101–114, <https://doi.org/10.1080/1573062x.2020.1857797>.
- [97] S. Yagi, S. Shiba, Application of genetic algorithms and fuzzy control to a combined sewer pumping station, *Water Sci. Technol.* 39 (1999) 217–224, [https://doi.org/10.1016/S0273-1223\(99\)00236-X](https://doi.org/10.1016/S0273-1223(99)00236-X), <https://www.sciencedirect.com/science/article/pii/S027312239900236X>.
- [98] R. Zavala-Yoe, H.M. Iqbal, R.A. Ramirez-Mendoza, Understanding the evolution of pollutants via hierarchical complexity of space-time deterministic and stochastic dynamical systems, *Sci. Total Environ.* 710 (2020) 136245, <https://doi.org/10.1016/j.scitotenv.2019.136245>.
- [99] H. Zhong, Y. Yuan, L. Luo, J. Ye, M. Chen, C. Zhong, Water quality prediction of mbr based on machine learning: a novel dataset contribution analysis method, *Journal of Water Process Engineering* 50 (2022) 103296, <https://doi.org/10.1016/j.jwpe.2022.103296>.
- [100] M. Zhu, J. Wang, X. Yang, Y. Zhang, L. Zhang, H. Ren, B. Wu, L. Ye, A review of the application of machine learning in water quality evaluation, *Eco-Environment & Health* 1 (2022) 107–116, <https://doi.org/10.1016/j.eehl.2022.06.001>.